UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Spatial-temporal analysis of hospital admissions due to heart failure in Portugal

Mestrado em Bioestatística

Daniel Filipe Viriato Pereira

Trabalho de Projeto orientado por:
Professora Doutora Marília Cristina de Sousa Antunes

2015

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Spatial-temporal analysis of hospital admissions due to heart failure in Portugal

Mestrado em Bioestatística

Daniel Filipe Viriato Pereira

Trabalho de Projeto orientado por:
Professora Doutora Marília Cristina de Sousa Antunes

2015

ACKNOWLEDGMENTS

I would like to thank IASIST Portugal for having provided the database with the aggregated information of Diagnosis Related Groups (DRG) concerning the number of hospital admissions in Portugal, between 2003 and 2012.

I am grateful to Frederico Calado for having convinced me to enrol in this Masters.

Susana Silva for all the support and friendship.

I am very grateful to my colleague and friend Carla Oiko, for all the notes, teachings, hours of study in the library, courage and patience.

I would like to express my sincere gratitude to my advisor Professor Marília Antunes for the continuous support of this project, for her dedication, motivation, and knowledge.

Finally, I would like to thank the unconditional support of my parents and my girlfriend Joana, who have always been present.

ABSTRACT

Introduction: Heart failure (HF) is a major reason for hospital admissions (HA), with a high socio-economic impact. Therefore it is of utmost importance to understand how HA due to HF are evolving.

Objective: This study aimed to conduct a special-temporal analysis of the annual number of HA due to HF in Portugal.

Methods: Hospital admissions due to HF, between 2003 and 2012, were extracted from National Diagnosis-related group database. Demographic and socioeconomic data were collected per district, from Statistics Portugal. Exploratory analysis was conducted in order to characterize the spatial-temporal characteristics of Portuguese population in terms of hospital admissions, demographic and socio-economic factors. Generalized linear mixed-effects models (GLMM) were used to estimate the annual number of HA. Spatial heterogeneity was corrected by considering region-related independent variables (IV): proportion of population aged ≥65, average monthly income and hospital access. Random effects were considered for all IV.

Results: The fixed effect estimates indicate that, in average, the number of HA due to HF increase by 7% per year. An increase of 1% in the proportion of population aged ≥65 accounts for an increase of 8% in HA. The increase of 100€ in the monthly income represents an average decrease of 5.8% in HA. By its turn, 1 more hospital per 100,000 inhabitants accounts for an increase of 2% in HA. These changes are conditional to all the other IV remaining unchanged. Estimated random effects accounted for spatial heterogeneity by introducing corrections around the fixed effects. The fitted model was compared to a GLMM without random effects for the region-related IV and a fixed effects model. Mean absolute deviations (MAD), used to assess goodness of fit, were 34.7, 56.5 and 131, respectively. Graphical representation also demonstrated that our model fitted better. Predictive ability of the model was assessed by MAD of forecast for 2012 based on 2003-2011 data (MAD=80.4).

Conclusions: The results of this study gave us some valuable information about how external factors influence evolution of HA in Portugal. Although this approach produced interesting results, the predictive ability could be further improved by the inclusion of other region-related variables.

Key-words: heart failure, hospital admissions, disease mapping, special-temporal analysis

RESUMO

Introdução: A insuficiência cardíaca (IC) é uma das principais causas de hospitalização, com um elevado impacto socio-económico. Desta forma, é importante perceber de que forma está a evoluir o número de hospitalizações por IC.

Objetivo: Este estudo teve como objetivo proceder à análise espaço-temporal do número de hospitalizações por IC em Portugal.

Métodos: Os dados de internamentos por IC, entre 2003 e 2012, foram extraídos da base de dados nacional de Grupos de Diagnósticos Homogéneos. Os dados demográficos e socioeconómicos foram retirados do Instituto Nacional de Estatística. Procedeu-se à exploratória dos dados a fim de caracterizar as características espaço-temporais da população Português em termos de hospitalizações, fatores demográficos e socioeconómicos. Foram utilizados modelos lineares generalizados mistos (GLMM) para estimar o número anual de internamentos. A heterogeneidade espacial foi corrigida considerando variáveis independentes (VI) relacionadas com a região: proporção de população com idade ≥65, rendimento médio mensal e acesso aos cuidados hospitalares. Foram considerados efeitos aleatórios para todas estas variáveis.

Resultados: As estimativas dos efeitos fixos indicam que, em média, o número de internamentos por IC aumenta 7% por ano. Um aumento de 1% na proporção da população com idade ≥65 contribui para um aumento médio de 8% no número de internamentos. O aumento de 100 € na renda mensal representa uma diminuição média de 5,8% no número de hospitalizações. Por sua vez, mais um hospital por 100.000 habitantes representa um aumento de 2% no número de internamentos. Estas alterações estão condicionadas a que todas as outras VI permaneçam inalteradas. Os efeitos aleatórios estimados contribuem para a heterogeneidade espacial através da introdução de correções em torno dos efeitos fixos. O modelo ajustado foi comparado a um GLMM sem efeitos aleatórios para as VI relacionadas com a região e um modelo de efeitos fixos. Os valores do desvio absoluto médio (DAM) utilizado para avaliar qualidade de ajuste foram 34,7, 56,5 e 131, respetivamente. A representação gráfica também demonstrou que o nosso modelo apresenta o melhor ajuste. A capacidade preditiva do modelo foi avaliada através do DAM da previsão para 2012 com base em dados de 2003-2011 (DAM = 80.4).

Conclusões: Este trabalho deu origem a uma série de informações relevantes acerca da forma como os fatores externos influenciam a evolução dos internamentos por IC em Portugal. Apesar de esta abordagem ter produzido resultados satisfatórios, a capacidade preditiva dos modelos pode ser melhorada através da inclusão de outras variáveis relacionadas com a região.

Palavras-chave: insuficiência cardíaca, hospitalizações, *disease mapping*, análise espaço-temporal

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACC	*American College of Cardiology*
ACSS	*Central Administration of Health Systems*
AHA	*American Heart Association*
AMI	*Average monthly income*
CAR	*Conditional autoregressive*
CHF	*Chronic heart failure*
DRG	*Diagnostic Related Groups*
EES	*Encontro Luso-Galaico de Estatística em Ambiente e Ecologia*
EPICA	*EPidemiologia da Insuficiência Cardíaca e Aprendizagem*
ESC	*European Society of Cardiology*
FHS	*Framingham Heart Study*
GEE	*Generalized estimating equations*
GLM	*Generalized linear models*
GLMM	*Generalized linear mixed-effects models*
GP	*General practitioner*
HA	*Hospital Admissions*
HF	*Heart Failure*
ISPOR	*International Society for Pharmacoeconomics and Outcomes Research*
IV	*Independent variables*
LRT	*Likelihood ratio test*
MAD	*Mean absolute deviation*
ML	*Maximum likelihood*
sqrtMSE	*Square root of mean squared errors*

SUMMARY

1

INTRODUCTION

1.1. HEART FAILURE OVERVIEW

Heart failure (HF) is a complex syndrome that can result from any structural or functional cardiac disorder, impairing the ability of the heart to function as a pump to support physiological circulation [**1**,**2**].

It is characterized by symptoms such as breathlessness and fatigue, and signs such as fluid retention. Common causes of HF include coronary artery disease, hypertension, valvular heart disease, cardiomyopathies and diabetes [**1**].

The European Society of Cardiology (ESC) defines HF as a syndrome in which a patient suffers from signs, symptoms, and objective evidence of an abnormality of the structure or function of the heart at rest [**3**]. The American College of Cardiology (ACC)/American Heart Association (AHA) define HF as a complex clinical syndrome that can result from any structural or functional cardiac disorder that impairs the ability of the ventricle to fill with or eject blood.

There is no single diagnostic test for HF and diagnosis involves a combination of medical history assessment, physical examination, and laboratory testing [**4**].

1.2. THE BURDEN OF DISEASE

About 26 million adults worldwide are living with heart failure, leading some to describe it as a global pandemic [**5**,**6**]. In comparison, 32 million are living with cancer and 34 million with HIV/AIDS [**7**,**8**].

In economically developed countries, up to one person in five is expected to develop heart failure at some point in their life, and even more people will be affected as family members, friends or healthcare professionals [**9**].

HF is a growing health burden with a substantial increase in incidence over the past two decades, and an annual incidence that is two-fold higher for each decade of age [**10**]. The Framingham Heart Study (FHS) estimated that the HF incidence approaches 10 per 1,000 population in patients aged ≥65 years and at 40 years, the lifetime risk of developing HF for both men and women is 1 in 5 [**9**]. The most recent AHA statistics reported that there are 870,000 new cases of HF annually (male: 415,000 cases/year; female: 455,000 cases/year) [**11**].

Available data from a number of geographies indicate that HF is common in most part of the world (1-2% of the adult population) and that there is an increase in prevalence of HF with advancing age (approximately ≥ 10% among patients aged ≥ 70 years) [**12**,**13**,**14**,**15**]. Another contributing factor to these increasing numbers is the improvement in treating heart attacks and other cardiovascular diseases that damage or place an extra burden on the heart. More patients with these conditions are surviving now than did in the past, but those who survive are at high risk of going on to develop heart failure [**6**].

Due to the growing trend of HF, demands on healthcare services are predicted to increase dramatically over the next decade as patient numbers rise owing to ageing populations, detrimental lifestyle changes and improved survival of those who go on to develop heart failure as the final stage of another disease [**16**].

1.3. HOSPITAL ADMISSIONS DUE TO HF

Heart failure is the most common reason for hospital admission in people over 65 years of age, accounting for about 1–4% of all hospital admissions in economically developed countries [**16**].

This is likely to be an underestimate because heart failure may be recorded as a secondary diagnosis, or may even go unrecorded, especially in the large number of patients who have other cardiovascular diseases [**17**].

Across the globe, 17–45% of patients admitted to hospital with heart failure die within 1 year of admission and the majority die within 5 years of admission [**16**].

In recent years, survival rates for patients with heart failure have improved in many parts of the world, nevertheless, about 2–17% of individuals admitted to hospital with heart failure die while in hospital [**18**,**19**].

Elderly patients hospitalized with heart failure are mainly women [**20**]. Although a number of studies of heart failure patients have indicated that survival rates are better in females than in males, recent research has shown that the long-term prospects for women are not as good as previously thought [**21**].

Caring for patients with heart failure comes at a high economic cost and accounts for about 1–3% of total healthcare expenditure in North America, Western Europe and Latin America [**22**,**23**,**24**]. In comparison, the total global expenditure on all healthcare goods and services in 2010 was in the region of US$650 trillion [**25**,**26**].

1.4. HF IN PORTUGAL

Epidemiological data regarding HF incidence, prevalence and prognosis in the Portuguese population and the only known study is the EPICA project (*EPidemiologia da Insuficiência Cardíaca e Aprendizagem* - Epidemiology of Heart Failure and Learning). The main goal of this study was to estimate the prevalence of chronic heart failure (CHF) in the population resident in mainland Portugal in 1998, according to the ESC Guidelines for the diagnosis of CHF [**27**]. This was a cross-

sectional observational study based on subjects attending primary health care centres in the community [**27**].

The study sample amounted to 5434 eligible subjects, evaluated by 365 general practitioners (GPs). Of these, 551 patients with CHF were identified. The overall prevalence of CHF in Portugal was estimated in 4.36%, 4.33% in males, and 4.38% in females. Age-specific CHF prevalence was as follows: 1.36% in the 25–49 years old group, 2.93% in the 50–59 years old group, 7.63% in the 60–69 years old group, 12.67% in the 70–79 years old group and 16.14% in group over 80 years old [**27**].

This study demonstrated that the overall prevalence of CHF in Portugal is slightly higher than that of other European countries and increases sharply with aging [**27**].

Accordingly, it is estimated that more than 400,000 people suffering from CHF in Portugal, which represents 5% of the population over 25 years old [**27**].

1.5. DISEASE MAPPING

Maps provide an efficient method of demonstrating distributions of an event in space. They provide a spatial pattern with an accuracy which cannot be attained by description or statistics, making them a powerful research tool. Maps help to record observations in a succinct format, outline analysis, formulate ideas and hypotheses and communicate findings [**28**].

In *Disease Mapping*, the term *disease* refers to the geographical distribution of a disease/event within a population, while *mapping* refers to the visual representation of the geographical distribution [**28**].

One of the most famous uses of disease mapping was the studies by John Snow of the cholera epidemics in London during the mid-19th century. Through a careful observation of his patients and by plotting where the case live, Snow showed that cholera could be spread through contaminated water supply. He designed a “dot map” of the residences of victims of the 1854 cholera epidemic in London, demonstrating a distinct cluster of cases around the water pump in Broad Street

(Figure 1). Later, investigations indicated that the pump had become contaminated by faecal material from a case of cholera.

Figure 1. Dot map of deaths from cholera in London (the arrow points to the Broad Street pump). Adapted from Lawson AB, *et al.* 2001 **[**28**]**.

Besides the visual information provided by maps, it is important to consider how many cases are expected to be found in a mapped area, since the distribution of the disease occurred within a population which itself has a spatial distribution. Moreover, the population has a demographic and socio-economic structure that has a spatial expression. To be able to assess whether any particular pattern of disease has arisen by chance, knowledge is required about the pattern that could arise from the underlying population [**28**].

The ultimate aim in disease mapping studies is the quantification of the deviation from the background level of the disease/event expected for the population of interest [**28**].

Many uses of disease mapping, such as the identification of spatial heterogeneity of disease risk or cluster investigation, are frequently constrained to a single time period, but data in epidemiology and public health are often available for time windows of several years. For those situations, it is possible to consider the analysis of disease maps with a temporal dimension. The most common format for observations is counts of cases of disease/event within small areas that are available for a sequence of time periods (spatial-temporal models) [**29**,**30**].

1.6. RESEARCH QUESTION

Heart failure has a high social and economic impact worldwide. Despite the existence of several therapeutic alternatives for the treatment of HF, it is not clear how this disease has evolved. International studies indicate that hospitalization due to HF are increasing, which can be strongly influenced by environmental, economic and social factors [**16**].

The present work aimed to study how the number of hospital admissions due to HF has evolved in the different districts of mainland Portugal.

2

MAIN GOALS

The main goal of this study was to assess the evolution of hospital admissions due to heart failure, in the different districts of mainland Portugal, between 2003 and 2012.

This study also assessed how demographic and socio-economic factors, featuring the different regions, influence the number of hospital admissions.

The modelling approach consisted in a Bayesian spatial-temporal analysis and generalized linear mixed-effects models (GLMM).

Work plan overview:

I) Literature review: to identify the main national and international studies about this theme

II) Acquisition and review of the statistical methodologies needed, namely, Disease Mapping, Bayesian spatial-temporal models, GEE and GLMM.

III) Conduct an exploratory analysis (tables, charts and maps) in order to understand data patterns and select the variables potentially associated with the response variable.

IV) Building of statistical models, its interpretation and validation.

V) Writing of thesis report.

The preliminary results of this project were presented in an oral communication at the congress "*I Encontro Luso-Galaico de Estatística em Ambiente e Ecologia (EES)"*, held 6-8 November 2014 in Vila Real, *Portugal* (Appendix 1).

The final modelling results were accepted for poster presentation at the *"International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 18th Annual European Congress"* to be held 7-11 November 2015 in Milan, Italy (Appendix 2).

3

METHODS

3. PROJECT OVERVIEW

This study was based on a dataset of hospital admissions per district of mainland Portugal. Each of the eighteen Portuguese districts contains data corresponding to a 10 years' time horizon, from 2003 to 2012 (Figure 2). The demographic and socioeconomic data used to characterize the population was also collected by district, for the same time period.

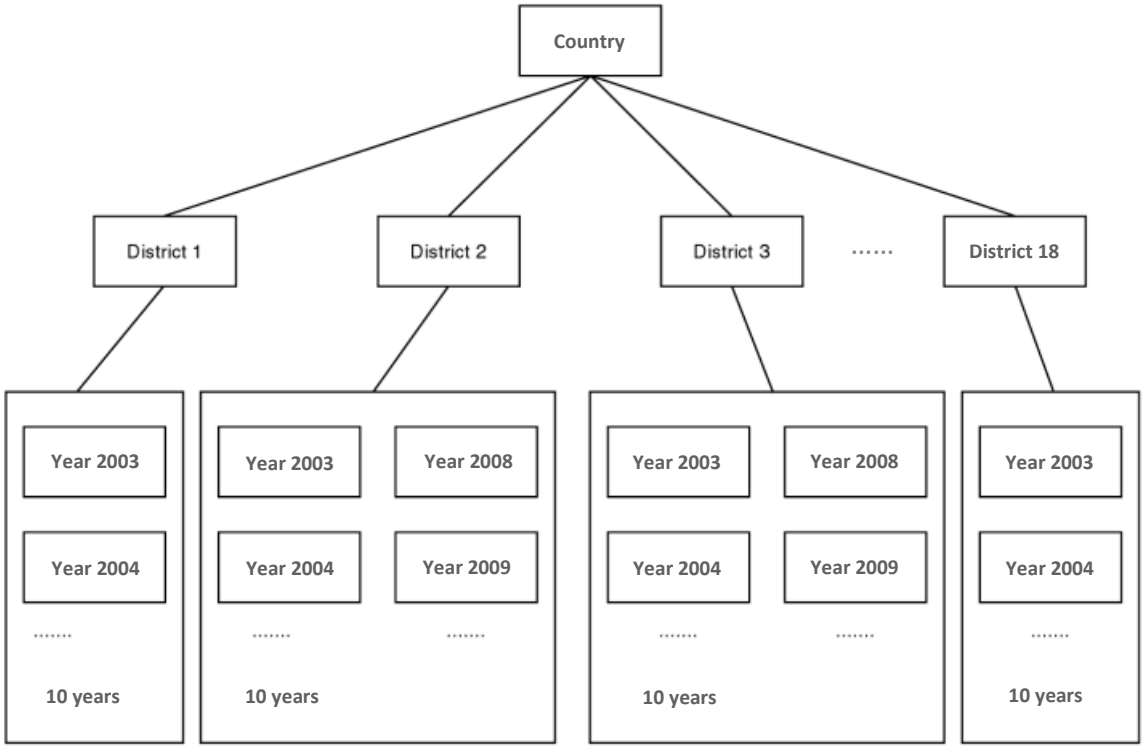
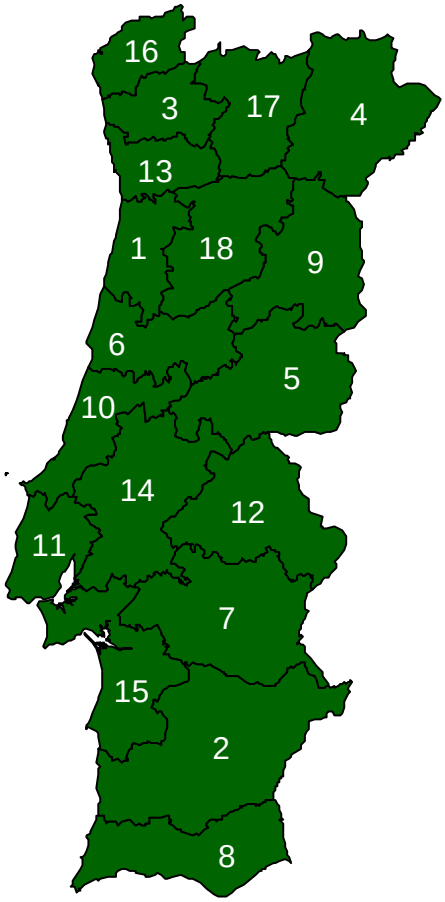


Figure 2. Dataset diagram of hospital admissions due to HF in Portugal.

In this work, the evolution of population patterns concerning the number of hospital admissions and other parameters will be explored in a temporal perspective, according to the spatial distribution of districts across mainland Portugal (Figure 3).



Districts of mainland Portugal:

1. Aveiro
2. Beja
3. Braga
4. Bragança
5. Castelo Branco
6. Coimbra
7. Évora
8. Faro
9. Guarda
10. Leiria
11. Lisboa
12. Portalegre
13. Porto
14. Santarém
15. Setúbal
16. Viana do Castelo
17. Vila Real
18. Viseu

Figure 3. Mainland Portugal map divided into districts.

3.1. VARIABLES & DATA SOURCES

Hospital admissions due to HF were extracted from the national database of Diagnostic Related Groups (DRG) of the Central Administration of Health Systems (ACSS) [**31**].

The number of hospital admissions was calculated based on the episodes of "hospital discharges" of public hospitals on the mainland Portugal, aggregated by district, on a 10 years' time horizon (2003-2012).

There were considered episodes with heart failure as the main diagnosis, including episodes with a length of 1 to 180 days. Episodes with more than 180 days were considered outliers and hence excluded (criterion defined by database owners). An overall of 110,520 heart failure episodes (average of 11,052 episodes per year) were considered in the analysis.

Demographic, social and economic data were collected from Statistics Portugal [**32**,**33**,**34**,**35**].

As demographic data, we considered the number of inhabitants by age and gender. The economic data includes the average monthly income of the population. In order to characterize the access to healthcare services we considered as social factors the number of hospitals and primary care centres (Figure 4).

All these covariates were collected per district of Portugal mainland for the same time period (2003 to 2012) that the hospital admissions data.

In the case of average monthly income, data was only available from 2004 to 2012. For that reason, the values for 2003 were estimated through linear regression models on time for each district, (Appendix 12).

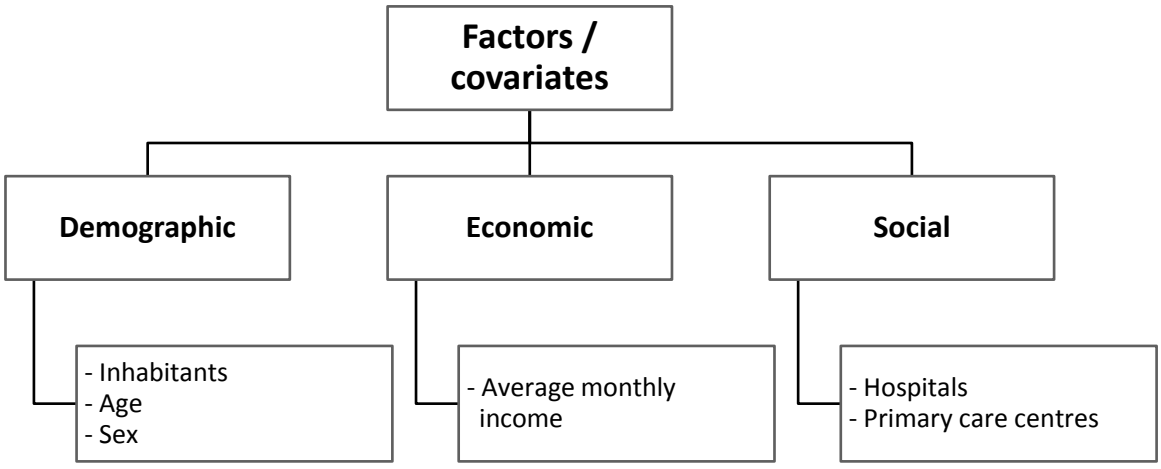


Figure 4. Covariates considered in the analysis.

A summary of the study variables and respective data sources is presented in Table 1.

Table 1. List of variables and data sources.

Variables	Data sources
Hospital admissions due to HF	Dados Nacional dos Grupos de Diagnóstico Homogéneos (GDH) da Administração Central de Sistemas de Saúde (ACSS) de 2003 a 2012 [**31**]
Population by age and sex	População média anual residente (N.º) por Local de residência (Distrito/ Região), Sexo e Grupo etário (Por ciclos de vida); Anual - Instituto Nacional de Estatística (INE), Estimativas Anuais da População Residente [**32**]
Average monthly income	Ganho médio mensal (€) por Localização geográfica (NUTS - 2002); Anual (1) – Instituto Nacional de Estatística (INE) [**33**]
Number of hospitals	Hospitais (N.º) por Localização geográfica e Natureza institucional; Anual - Instituto Nacional de Estatística (INE), Inquérito aos Hospitais [**34**]
Number of primary care centres	Centros de saúde (N.º) por Localização geográfica e Tipo de serviço; Anual - Instituto Nacional de Estatística (INE), Inquérito aos Centros de Saúde [**35**]

3.2. EXPLORATORY ANALISIS

A series of exploratory analysis were conducted using R Statistics v3.2.0® and Microsoft Excel 2010® softwares in order to characterize the Portuguese population per district and overtime in terms of hospital admissions, demographic and socio-economic factors.

Disease Mapping techniques were used in order to study the patterns of hospitalization rates due to HF, demographic and social-economic changes (age, sex, average monthly income, number of hospitals and number of primary care centres).

The analysis included line plots, scatterplots, boxplots, lattice xyplots, maps and descriptive tables.

The exploratory analysis was important to:

- Describe the data;
- Study the spatial and temporal evolution of hospital admissions and region-related covariates;
- Select the relevant covariates to include in the modelling analysis.

3.3. STATISTICAL MODELS

The Modelling Approach

In research areas such as medicine or public health, data often exhibit a structure which is inherent to the units of the population where data come from. This structure is frequently nested (also called hierarchical) with the statistical units being observed at higher level units. Very often, the data is not observed at the individual level but refers to counts by area, e.g. regions of a country. These are usually called small-area data [**29**].

Statistical methods employed in the analysis of small-area data are diverse and range from basic exploratory and descriptive methodology to particular spatial

statistical methods. The basic characteristic of such data is its discrete nature - counts of events (e.g. hospital admissions) within defined geographical regions [**36**].

3.3.1. Spatial-temporal models

In most of the studies, the region of interest has been divided into n contiguous subregions covering the totality of the region. Let y_i represent the number of cases of the event of interest for each area $i = 1,\ldots, n$. Generally, in the literature, it is assumed that y_i follows a Poisson distribution with mean $e_i r_i$, where e_i is the expected counting of the event of interest and r_i represents the relative risk, i.e.,

$$y_i \sim Poisson\ (e_i r_i).$$

The expected counts are a known quantity based on common risk factors, and play an important role in standardizing the information from areas of different characteristics. Therefore, the unknown quantities of interest are the relative risks, r_i.
To estimate the disease rates based on the maximum likelihood (ML) estimate of r_i, that is, to estimate of r_i, by y_i/e_i, the crude rates, is misleading when the event of interest is rare and/or the areas are small [**37**]. Also, the ML estimates are not able to accommodate the possible spatial dependence. The homogeneity within each area is another aspect to be taken into account as well as heterogeneity between regions. For that reason, some variation in the model should be included so that the estimated relative risks reflect such aspects [**36**,**38**].

A classical reference in the literature on modelling relative risks is Besag *et al.* 1991 [**39**] who assumes that, for area i, the log relative risks are modelled through

$$\log(r_i) = \alpha + \beta x_i + b_i + u_i,$$

where α is a common intercept for the entire region, x_i is a vector of covariates and u_i and b_i are random effect terms for the subregion i. Random effect term u_i is such that $u_i \sim N(0, \sigma_u^2)$, $i = 1,\ldots,n$, representing some unstructured noise term, and b_i accounts for the spatial dependence,

$$b_i | b_j, j \neq i \sim N\left(\frac{\sum_{j\in\delta_i} w_{ij} b_j}{\sum_{j\in\delta_i} w_{ij}}, \frac{\sigma_b^2}{\sum_{j\in\delta_i} w_{ij}}\right), i = 1, \ldots, n,$$

where δ_i is the set containing areas adjacent to *i*, and W_{ij} is the weight that neighbouring area *j* has on *i*. We will consider the simplest, and most frequently used, neighbouring structure which is based on adjacent areas, where regions are considered neighbours if they share a common boundary. This brings further simplification since in this case W_{ij} = 1 if *i* and *j* share boundaries and W_{ij} = 0 otherwise [**38**,**40**].

The prior distribution above is known as conditional autoregressive (CAR) prior. We will denote it by $CAR(\sigma_b^2)$.

Many small-area datasets are collected over time, enhancing the need of extending the models to the spatial-temporal case. This is a straightforward approach in which space is indexed by region (e.g. district) and time indexed by year.

In our case, data concerns the number of hospital admissions due to HF (*Y*) between 2003 and 2012 (*t* = 1, …, 10) in 18 districts of mainland Portugal (*i* = 1, …, 18).

Let Y_{it} be the number of hospital admissions in region *i* at year *t*. Then,

$$Y_{it} \sim Poisson\ (\mu_{it}),$$

with $E(Y_{it}) = \mu_{it} = e_{it}\lambda_{it}$, where e_{it} is the offset term, which represents the expected number of cases for region i at time t and λ_{it} represents the relative risk of the event for region i at time t [**36**,**38**].

Although the interest lays in the number of hospital admissions due to HF, it is important to notice that this number is highly dependent on the number of individuals exposed to the risk of the event of interest. Such issue is overcome by modelling the relative risk of the event.

A common approach to estimate e_{it} is to consider that the incidence rate of the event of interest is constant over time and space and hence

$$e_{it} = population_{it} \times \frac{total\ number\ of\ cases}{total\ exposed\ population}.$$

The total number of cases corresponds to the total number of hospital admissions due to HF in the 18 regions over the 10 years and the total exposed population

corresponds to the number of person-years lived by the population of the 18 regions over the 10 years. This latter value is approximated by the sum of the total population of the country for the 10 years. $Population_{it}$ corresponds to the population of each region *i* at year *t* [**36**,**38**].

The logarithm of the expected number of cases can then be expressed as

$$\log(\mu_{it}) = \log(e_{it}) + \log(\lambda_{it})$$

and, by its turn, the logarithm of λ_{it}, the log relative risk, is expressed as a linear function of time, *t*, and the covariates.

Most of the studies on disease mapping do not deal with the spatial and temporal components jointly. The presence of longitudinal information and spatially referenced data for disease incidence, encourages the study and development of new models to deal with this important issue.

A simple approach consists in modelling the log relative risk through

$$\log(\lambda_{it}) = \alpha + \beta_i x_{it} + b_i t + u_i,$$

where the random effect term u_i is modelled as a CAR(σ_u^2) [**38**].

This model states that the observed number of hospital admissions in a particular region in a certain period of time follows a Poisson distribution whose parameter is a function of the expected number of hospital admissions, the year and district and a set of covariates represented by *x*. In this model it is assumed that the log relative risk changes linearly in time [**38**,**40**].

Because small-area data contain the influence of variables affecting the local populations which are not accounted for in standardized rates, the covariates in vector *x* are usually socio-demographic variables. For heart failure, we expect the number of hospital admissions to be influenced by demographic and socio-economic factors [**29**,**30**].

These models are estimated within a Bayesian framework. The prior distribution for the parameter is Normal with zero mean and small value for the precision parameter,

e.g. N(0,0.0001). For the parameters β_i and b_i the prior distributions are also zero mean Normal, with precisions τ_β and τ_b, both distributed as Gamma, for example, Gamma(0.1,0.0001).

Due to difficulties inherent to the estimation of the above described models, an alternative simpler approach to model the random effect term u_i is to consider these terms independent and normally distributed,

$$u_i \sim N(0, \sigma_u^2).$$

This change implies that it is no longer considered the possibility of spatial dependence in the error structure of the model. In this approach, which falls in the class of Generalized Linear Mixed Models, the inclusion of random effects should account for all the differences between the districts [**38**].

3.3.2. Generalized Linear Mixed Models

Generalized linear models (GLM) are a class of fixed effects regression models which include linear regression, logistic regression, and Poisson regression.
In a GLM, the specifications concern (1) the linear predictor, denoted as η_i, of the form $\eta_i = x_i'\beta$, where x_i is the vector of independent variables for subject i with fixed effects β; (2) the link function, g, which converts the expected value μ_i of the outcome variable Y_i to the linear predictor η_i , $g(\mu_i) = \eta_i$; and (3) the form of the variance in terms of the mean μ_i. The link function and the variance depend on the distribution of the dependent variable, Y_i , which is assumed to fall within the exponential family.
In the fixed effects models, it is assumed that all observations are independent of each other. Hence, these models are not suitable for data exhibiting dependence, in particular clustered and/or longitudinal data. Clustered data concerns designs where subjects are observed nested within units such as, for example, schools or hospitals. In longitudinal designs, repeated measurements are taken for each individual, resulting in observations nested within individuals. These models are also known as Multilevel Models due to the hierarchical structure of the data, in which the first level

observations (subjects or repeated observations) are nested within the higher second level observations (clusters or subjects), and so on [**41**,**42**].
The analysis of multilevel data should be able to account for the correlation in the data. For that purpose, cluster and/or subject random effects can be added into the regression model, resulting in a mixed (fixed + random) effects model.

Our data does not fall exactly in the above described cases since we do not have individual observations (subjects) nested into clusters nor repeated observations by subject. Still, these models are suitable since our data is aggregated by region and collected over time, resulting in temporal areal data nested into districts (see Figure 2).

Mixed models for continuous normal outcomes have been deeply developed and extensive literature and software can be found. For non-normal data, there have also been many developments, which fall under the family of generalized linear mixed models (GLMM), which extend GLM by the inclusion of random effects in the predictor [**42**].
The modelling approach is the same for longitudinal and clustered data and serves our hierarchical data as well. For simplicity the notation will be adapted to our case. Let i denote the level-2 units (the districts) and let j denote the level-1 units (the observations over time).
Assume there are i = 1, . . .,N districts and j = 1, . . . , n_i observations nested within each district. A random-intercept model is the simplest mixed model, augmenting the linear predictor by adding a single random effect for district i,

$$\eta_{ij} = x'_{ij}\beta + b_i,$$

where b_i is the random effect (one for each district). These random effects represent the influence of district i on its repeated observations that is not captured by the observed covariates. When dealing with subjects instead of regions, these are treated as random effects because the sampled subjects are thought to represent a population of subjects, and they are usually assumed to be distributed as N(0, σ_b^2) [**41**]. In our case, the districts were not sampled from a “population” of districts of the country but the object of interest (the number of hospital admissions due to HF) is heterogeneous between districts indicating clearly that the units should not be treated

as homoscedastic. The parameter σ_b^2 indicates the variance in the population distribution, and therefore the degree of heterogeneity of districts. Including the random effects, the expected value of the outcome variable is given as

$$\mu_{ij} = E\left[Y_{ij}|b_i, x_{ij}\right],$$

which is related to the linear predictor via the link function. This is the expectation of the conditional distribution of the outcome given the random effects. This is why GLMM are often referred to as conditional models in contrast to the marginal generalized estimating equations (GEE) models which represent an alternative generalization of GLM for correlated data [**41**].

The model can be easily extended to include multiple random effects. In our case it would make sense to have a random district intercept and a random linear time-trend. Let z_{ij} be the r × 1 vector of variables having random effects (with a column of ones included for the random intercept). The vector of random effects b_i is assumed to follow a multivariate normal distribution with mean vector 0 and variance–covariance matrix Σ_b.

The model is now written as

$$\eta_{ij} = x'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i.$$

In our case, because the outcome variable corresponds to a count, that is, if Y_{ij} corresponds to the value of the count variable associated with district I and time point j, assuming this count to be drawn from a Poisson distribution, then the mixed Poisson regression model indicates the expected number of counts as

$$\log \mu_{ij} = \eta_{ij}$$

with the linear predictor $\eta_{ij} = x'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i$. In the general case for longitudinal data, often subjects (individuals) are not followed for the same time interval and hence they are not equally exposed to the risk of occurrence of the event of interest (e.g. number of asthma episodes) [**41**]. In our case, we are interested in the number of hospital admissions per year and per district. Since the districts are not equally populated, the number of individual at risk of suffering the event of interest varies from district to district. For this, let e_{ij} represent the expected number of hospital admissions due to HF in district i and year j, as explained above in this section. The linear predictor is now augmented as

$$\eta_{ij} = \log e_{ij} + x'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i.$$

which can also be expressed as

$$\mu_{ij} = e_{ij} \times \exp\left(x'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i\right),$$

or

$$\frac{\mu_{ij}}{e_{ij}} = \exp(x'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i)$$

to reflect that it is the relative risk per district that is being modelled. As referred above in this section, e_{ij} can be calculated (estimated form the data) assuming that rate of hospital admissions due to HF is constant over space (the districts) and time (the years of study). The term $\log(e_{ij})$ is often called an offset.

Assuming the Poisson process for the count Y_{ij} , the probability that $Y_{ij} = y$, conditional on the random effects **b**, is given as

$$P\left(Y_{ij} = y|b_i, x_{ij}, z_{ij}\right) = exp\left(-\mu_{ij}\right)\frac{u_{ij}^{y}}{y!}.$$

Departures from the distribution assumption can occur, with the excess of zeroes or overdispersion being the most common cases. In disease mapping problems, the latter are more frequently found. Alternative models include generalized Poisson and negative binomial. Although fairly well studied, such models are not easily found implemented and available for use [**41**].

Parameter estimation in GLMM typically involves maximum likelihood (ML) or variants of ML. Additionally, the solutions are usually iterative and numerically very intensive. Several approaches and methods have been proposed (Gauss-Hermite, Laplace approximation, quasi-likelihood, MCMC) with a lot of discussion around this subject [**43**].

In R there are some packages available for estimating GLMM [**44**]. Our choice fell on lme4 (using Laplace approximation) as it contains implemented features as Wald test (`summary`), likelihood ratio test (LRT) (`anova`) for model selection and `predict` for the calculation of fitted values and performing prediction using new data [**42**,**45**].

The model selection strategy is not straight forward, starting from the simple fact that the usual linear regression's sums of squares terms and degrees of freedom for the numerator and denominator of an F test cannot be used. All the issues that arise with regular linear or generalized-linear modelling such as inadequacy of p-values alone, the need to understand how models are parameterized, the dangers of overfitting

among others also apply, but more severely, to mixed models. Guidance on model selection, interpretation and prediction is not easily found, especially when departing from the most common applications of the models.

The use of LRT is an adequate tool for comparison of nested models, in particular if the models differ by a fixed effect. Because the new parameter estimator is asymptotically normally distributed, the use of chi-square distribution to assess the magnitude of the difference between the log likelihood of the two models is adequate. When nested models differ by the inclusion of random effects, the new parameter is the variance of the added random effect. Because its distribution is not normal (and can be quite skewed), the result of LRT does not offer much confidence. Some approaches based on simulation procedures are discussed in the literature but its implementation and use is far from being easy [**45**].

4

RESULTS

4. EXPLORATORY ANALYSIS

An exploratory analysis of the data presented in the previous section was performed in order to understand data patterns and select the covariates potentially associated with the response variable (number of hospital admissions).

This exploratory analysis allowed the characterization of the population in terms of hospital admissions, demographic and socio-economic factors.

Data is presented in tables and graphics that show the evolution of variables over time and maps that provide a spatial pattern distribution of population parameters, in the different periods.

4.1. HOSPITAL ADMISSIONS

The number of hospital admissions due to HF in Portugal varied between 8,750 in 2003 and 13,479 in 2012, corresponding to an increase of 35% in a 10 years' time span, with an average annual increase of 5% (Figure 5).

Porto and Lisboa are the districts with the highest number of hospital admissions. From 2003 to 2008, Porto leaded this ranking, but from 2007 until 2010 it has a sharp decrease in the number of hospitalizations. On the other hand, the growing trend demonstrated by Lisbon, makes it the district with the highest number of hospital admissions between 2008 and 2012, reaching a maximum of 2,466 in 2012 (Figure 6).

In other districts, the number of hospitalizations remained stable, with minor fluctuations over time (Figure 7).

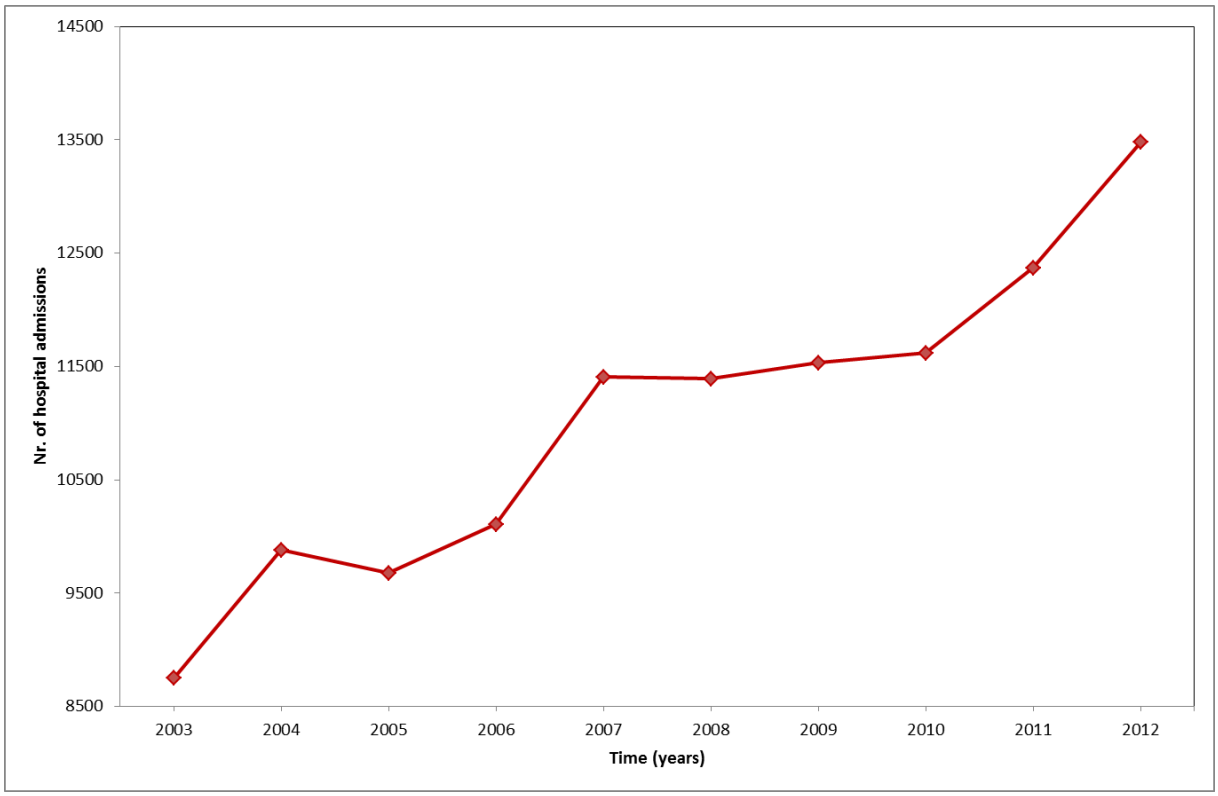


Figure 5. Number of hospital admissions due to HF in mainland Portugal, over time.

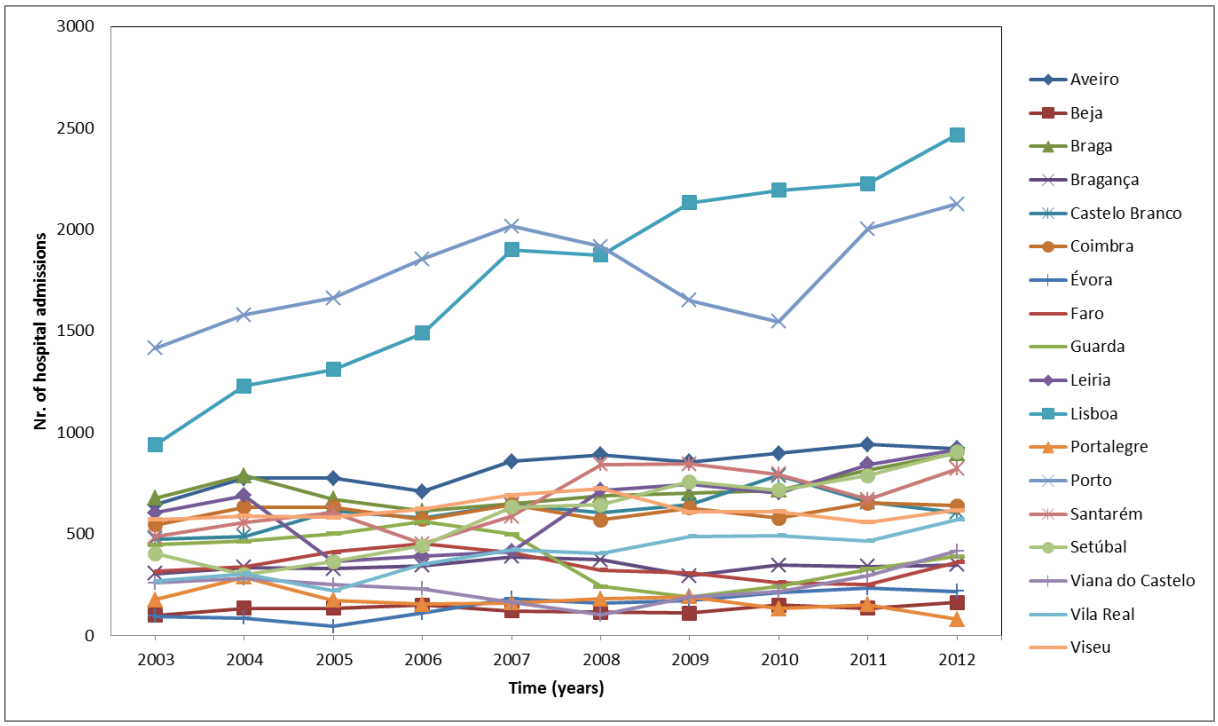


Figure 6. Number of hospital admissions due to HF per district of mainland Portugal, over time.

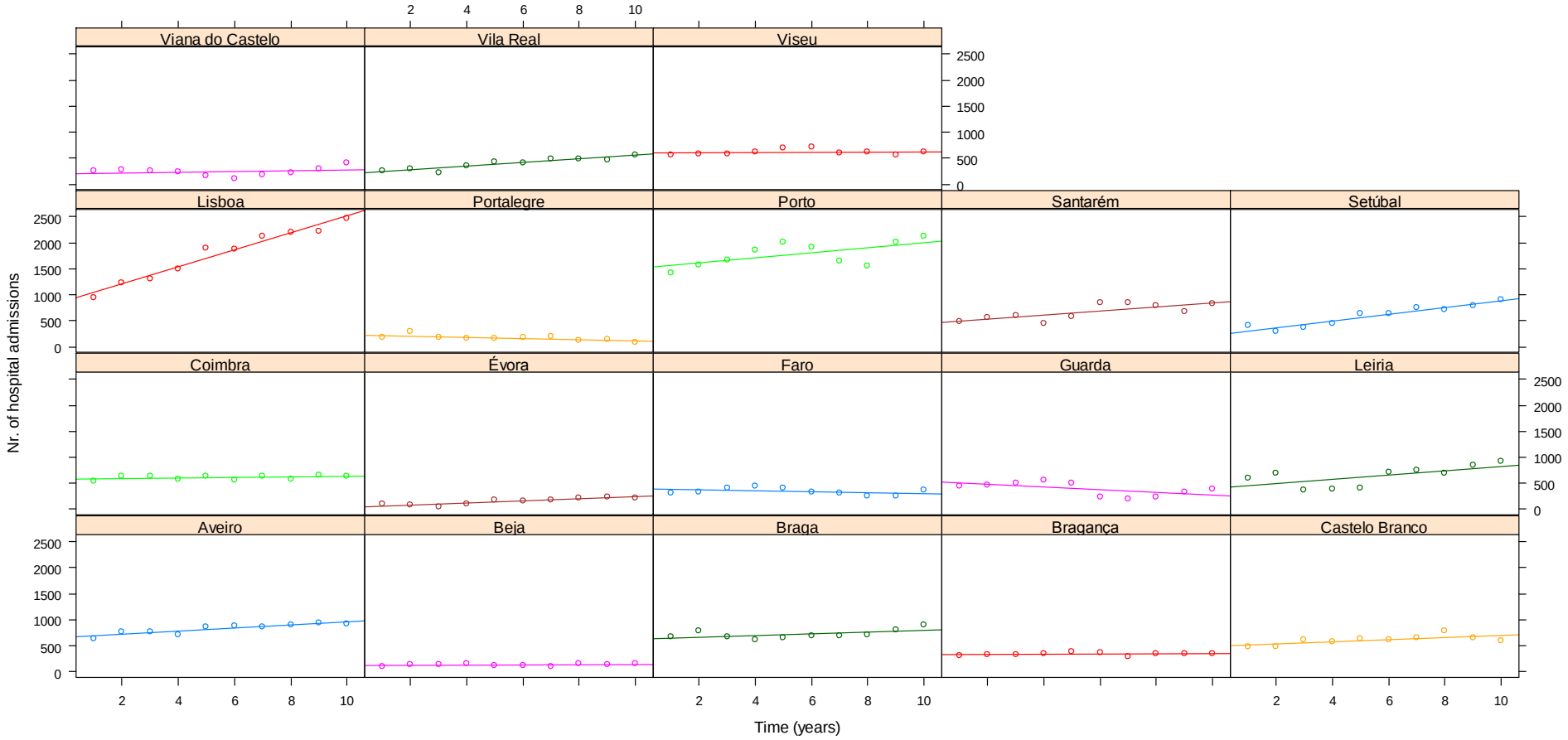
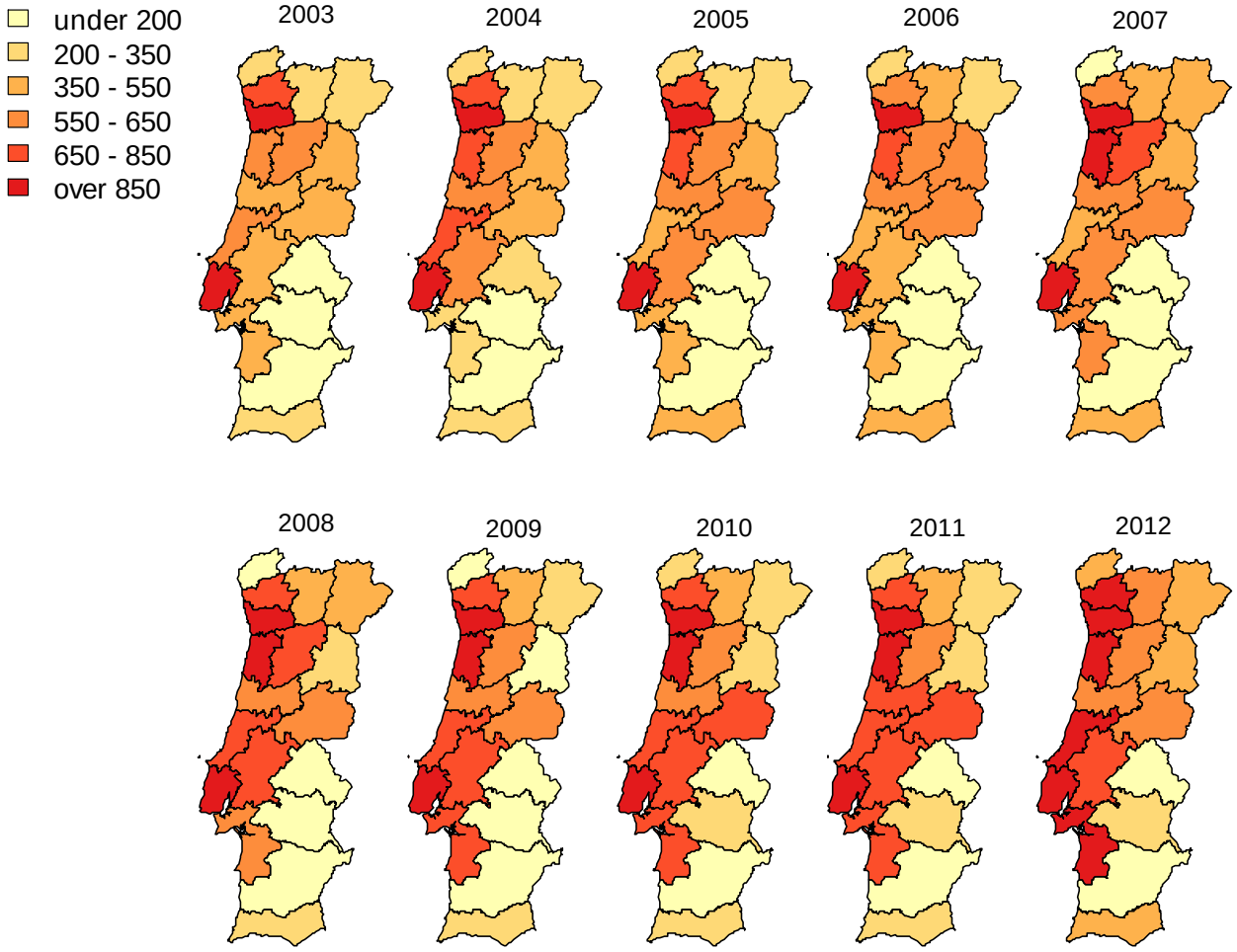


Figure 7. Evolution trend of hospital admissions due to HF per district of mainland Portugal.



According to the map, the districts with the highest number of hospital admissions are located in the west of mainland Portugal.

This is due to the fact that these districts are the most populated.

The overall number of hospital admissions increased throughout the national territory between 2003 and 2012.

In 2012, 6 out of the 18 districts (Braga, Porto, Aveiro, Leiria, Lisboa and Setúbal) exceeded the 850 hospitalizations (Figure 8).

Figure 8. Maps representing the number of hospital admissions due to HF per district of mainland Portugal, over time.

In mainland Portugal, the number of hospital admissions per 100,000 patients varies between 88 in 2003 and 135 in 2012 (Figure 9).

When the number of hospital admissions is adjusted per 100,000 inhabitants, the districts' pattern is different from the previous results. Despite the east has a lower number of hospital admissions, it is also less populated than the west, resulting in higher rate per 100,000 inhabitants.

In 2003, Guarda, Cartelo Branco and Bragança were the districts with the highest number of hospital admissions per 100,000 inhabitants. However, a sharp decrease (more than 50%) was observed in Guarda district between 2006 and 2009 (Figure 10).

Some fluctuations were observed in Castelo Branco and Brangança, but these remained the districts with the highest number of hospitalizations per 100,000 inhabitants

A growing trend was observed in Vila Real that in 2012 becomes the second district with the highest number of hospital admissions per 100,000 inhabitants.

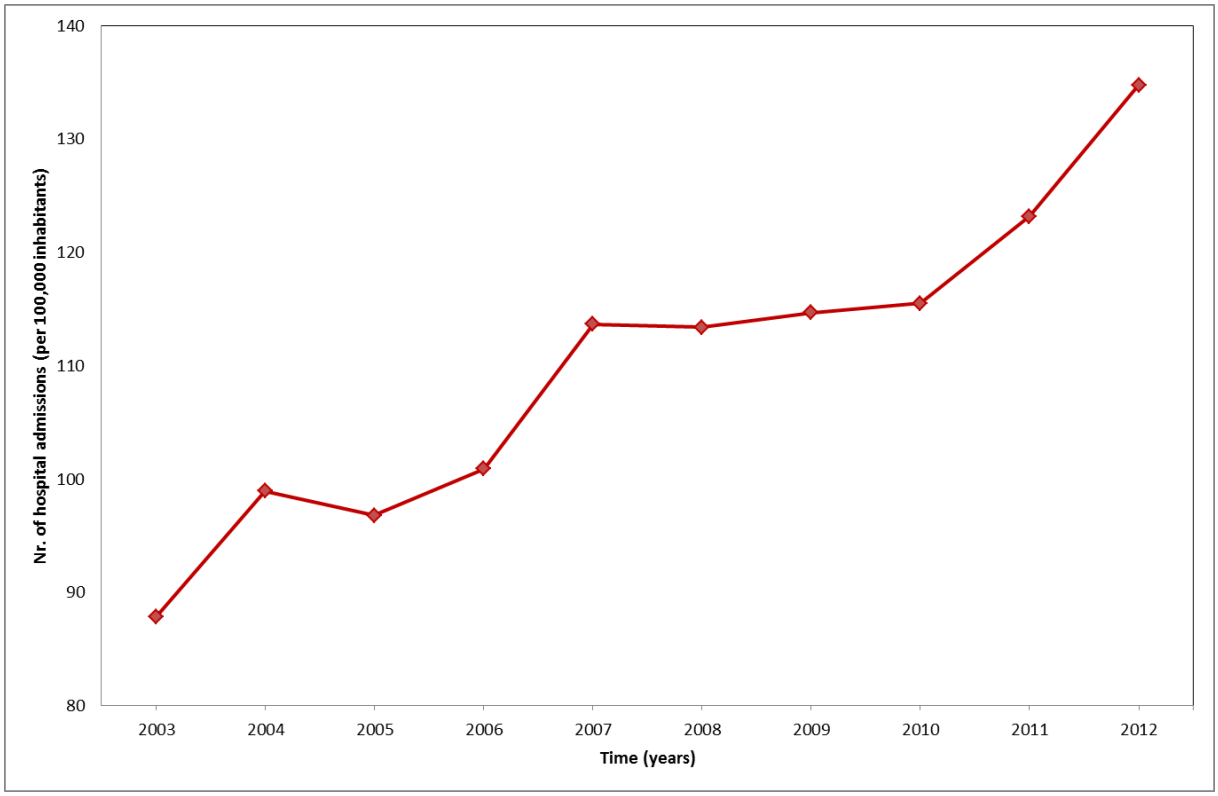


Figure 9. Number of hospital admissions due to HF per 100,000 inhabitants in mainland Portugal, over time.

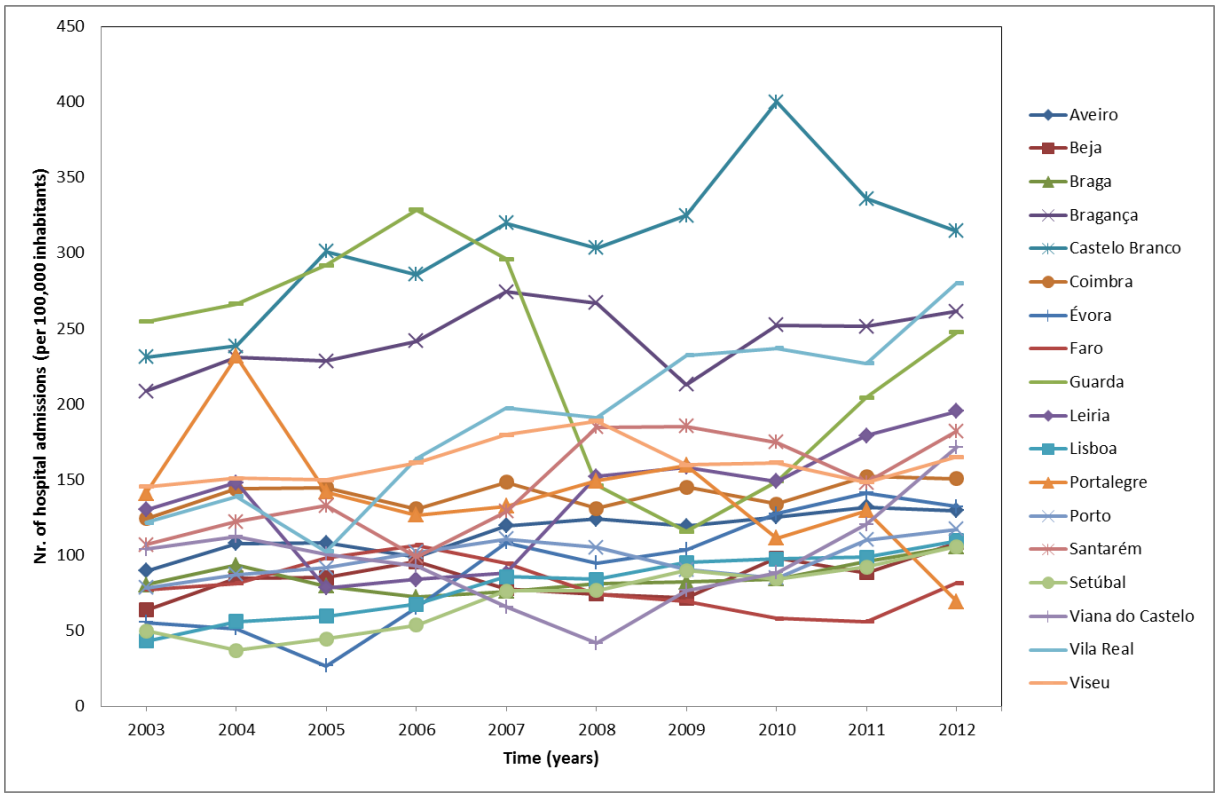


Figure 10. Number of hospital admissions due to HF per 100,000 inhabitants, per district of mainland Portugal, over time.

The evolution trends presented in Figure 11 show that 16 out of the 18 districts had an increase in the number of hospital admissions per 100,000 inhabitants, between 2003 and 2012. Only Guarda and Portalegre had a decreasing trend in the number of hospitalizations that could be related with the hospital referencing network.

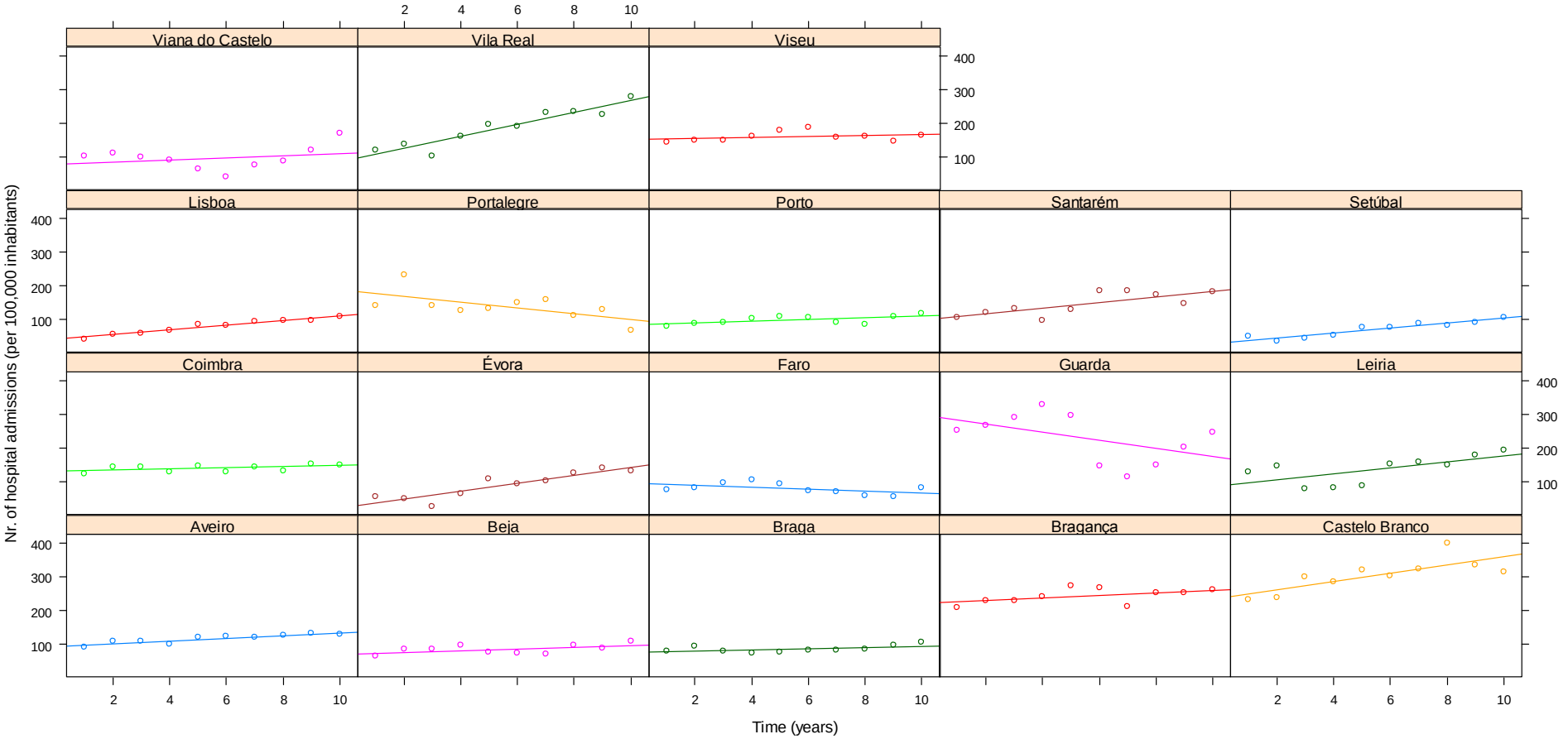


Figure 11. Evolution trend of hospital admissions due to HF per 100,000 inhabitants, per district of mainland Portugal.

The average number of hospital admissions increased from 87.83 in 2003 to 134.74 in 2012, with a 10 years average of 109.98 (Table 2).

Table 2. Average number of hospital admissions due to HF per 100,000 inhabitants per year.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	10 years
87.83	98.94	96.77	100.90	113.66	11.34	114.67	115.50	123.15	134.74	109.98

In the boxplot (Figure 12) it is possible to identify two outliers that influence the average number of hospital admissions per 100,000 inhabitants, over time: Guarda (between 2003 and 2006) and Castelo Branco (2005-2006 and 2009-2011), which supports the results previously described.

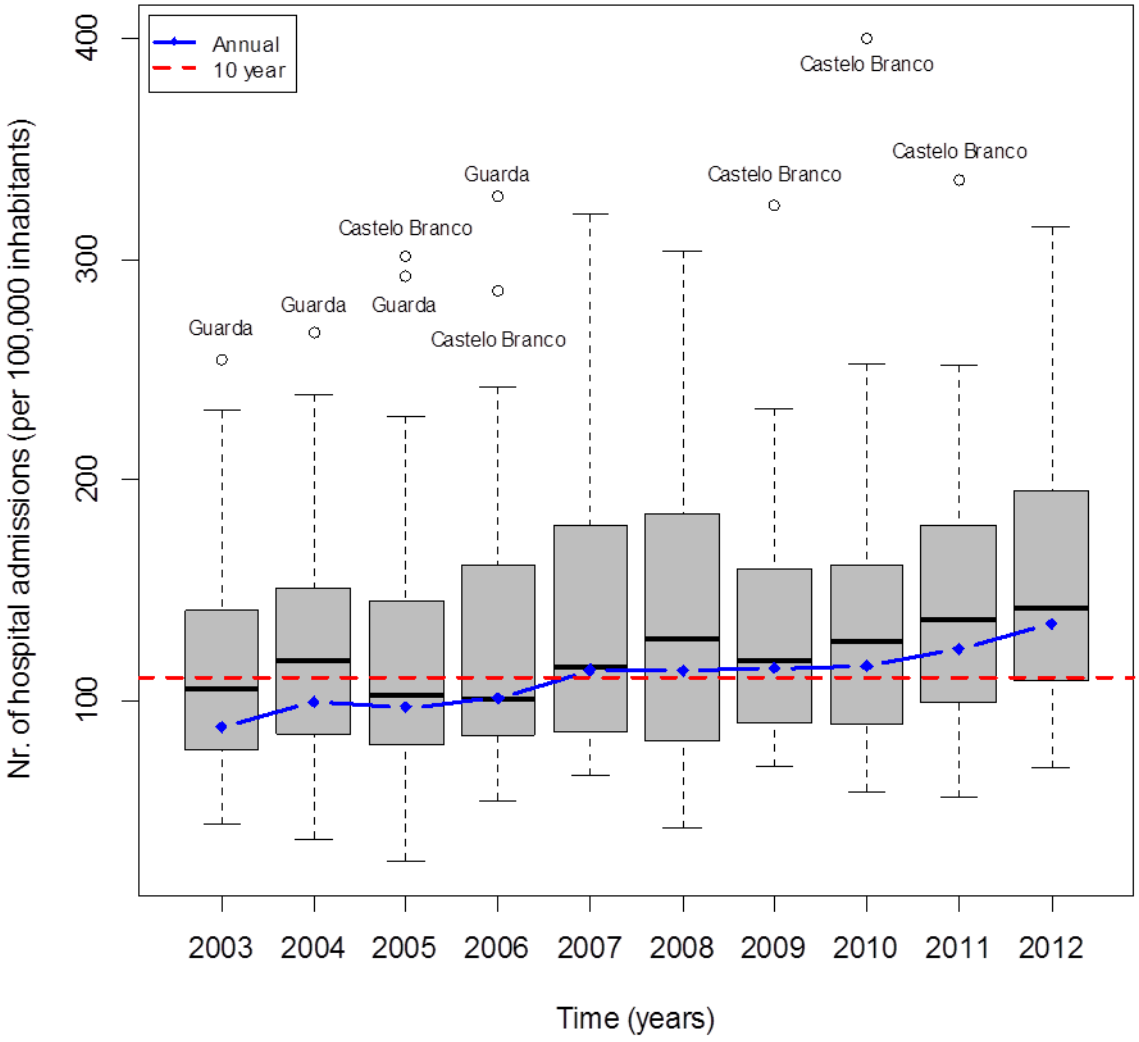
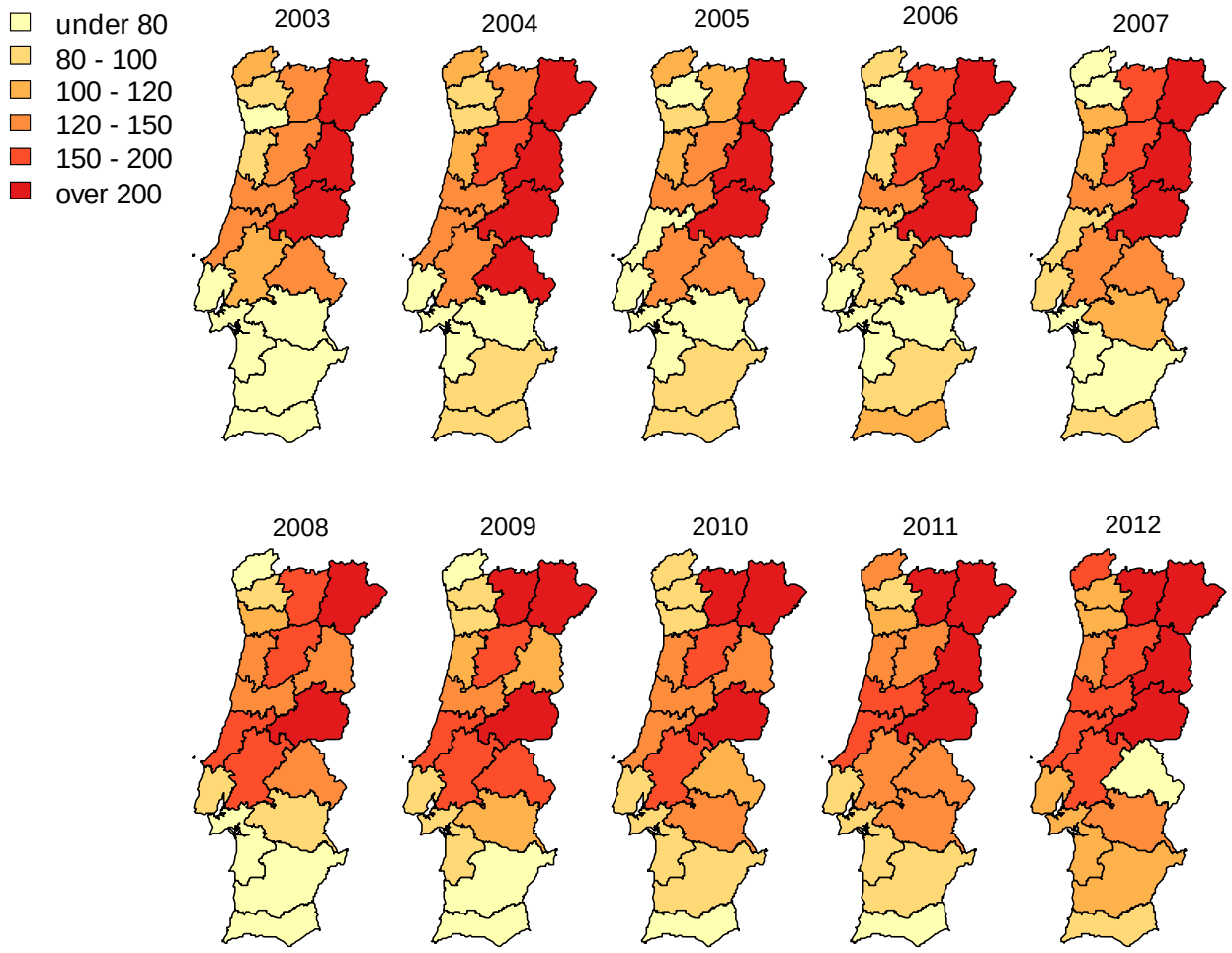


Figure 12. Boxplot with the number of hospital admissions due to HF per 100,000 inhabitants, per district of mainland Portugal, over time.



The map of the number of hospitalizations per 100,000 inhabitants has a different pattern than the previous one.

The districts with the highest number of hospital admissions per 100,000 inhabitants are located in the east of mainland Portugal, instead of west.

Since the incidence of HF is higher in older populations, this pattern could be explained by the aging in these districts.

The overall number of hospital admissions increased throughout the national territory between 2003 and 2012.

In 2011 and 2012, 4 out of the 18 districts (Vila Real, Bragança, Guarda and Castelo Branco) exceeded the 200 hospitalizations per 100,000 inhabitants.

Lisboa and Porto that previously presented the highest number of hospitalizations nationwide, between 2003 and 2012, presented now a rate that varies between 43 and 110, lower than most of the other districts.

Figure 13. Maps representing the number of hospital admissions due to HF per 100,000 inhabitants, per district of mainland Portugal, over time.

4.2. DEMOGRAPHIC FACTORS

Populations have demographic and socio-economic structures that could have spatial and temporal expression.

The following analysis describes the evolution of the number of inhabitants in Portugal per district and the influence of age and sex in the distribution patterns.

The Portuguese population remained approximately constant between 2003 and 2012, with a variation lower than 1%, being around 10 million inhabitants (Figure 14).

In order to identify small fluctuations in the growing trend of the Portuguese population, it was used a reduced scale in the chart of Figure 14. This allows to find a small decrease in the Portuguese population, since 2010 (-0.5%).

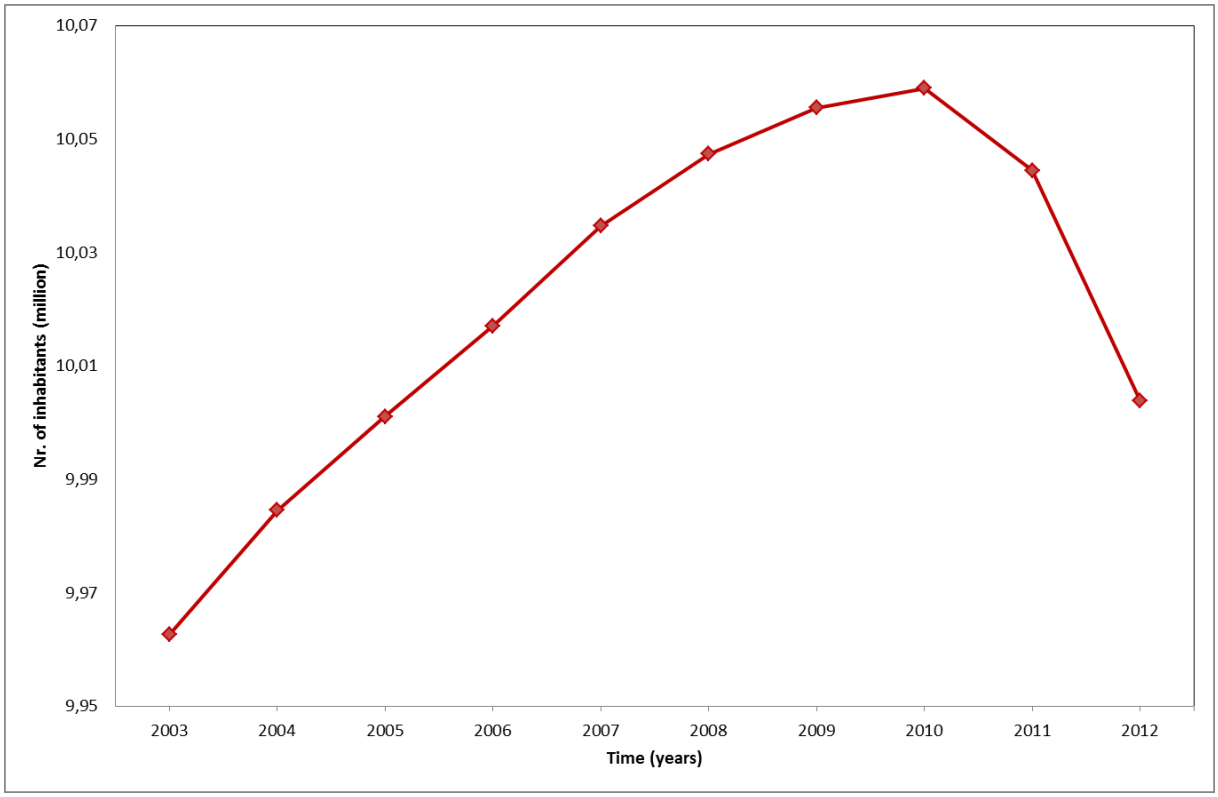


Figure 14. Evolution of resident population (in million inhabitants) in mainland Portugal, over time.

The most populated districts are located in the west of mainland Portugal: Lisboa (over 2 million inhabitants), Porto (over 1.5 million inhabitants), Braga, Setúbal and Aveiro (over 0.5 million inhabitants) (Figure 15).

Despite these districts also present the highest record of hospital admissions due to heart failure, the larger number of inhabitants (2 to 20 times more inhabitants than the others districts) explains the lower rate of hospitalizations per 100,000 inhabitants, when compared with the east districts, as previously described.

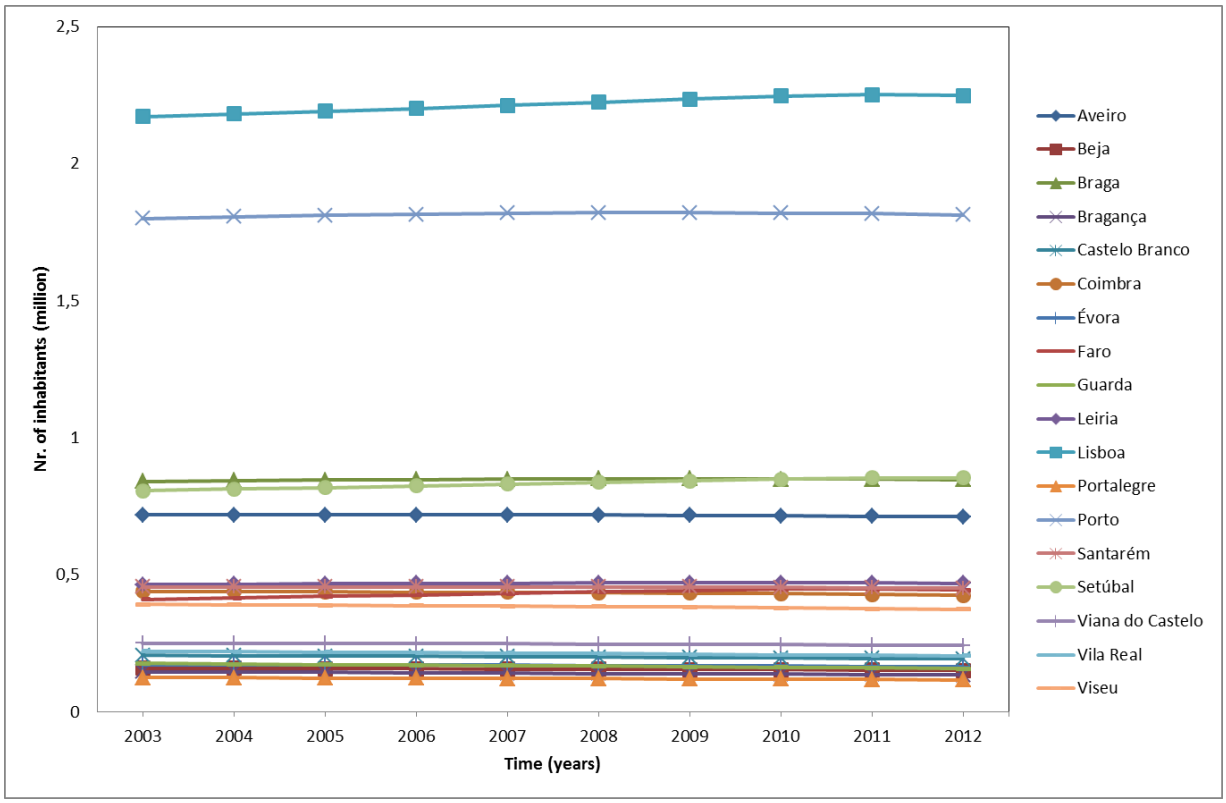


Figure 15. Evolution of resident population (in million inhabitants) per district of mainland Portugal, over time.

As the previous chart, the number of inhabitants per district remained flat over the time, with minor fluctuations. Only Guarda presented a decrease of 10% in the population (Figure 16).

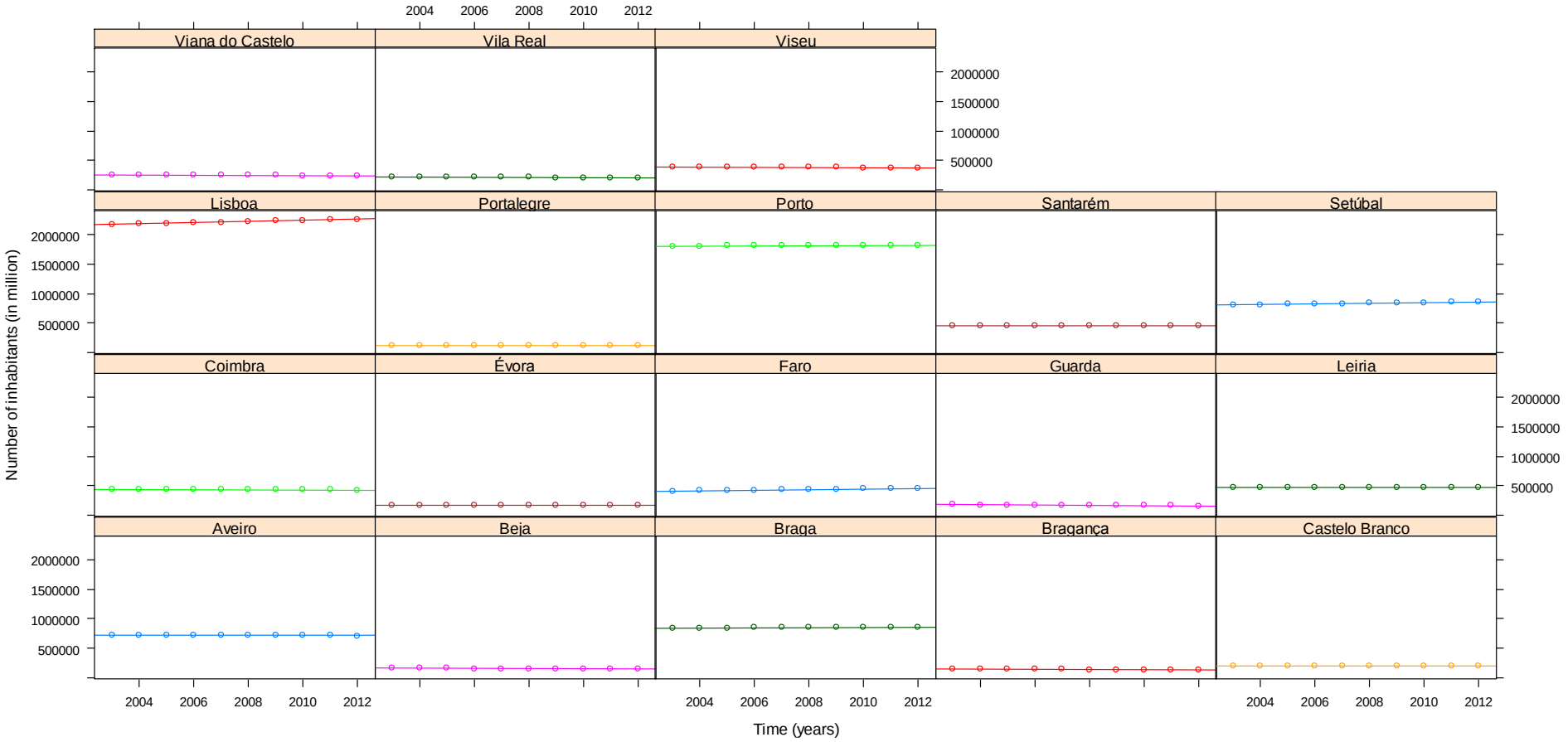
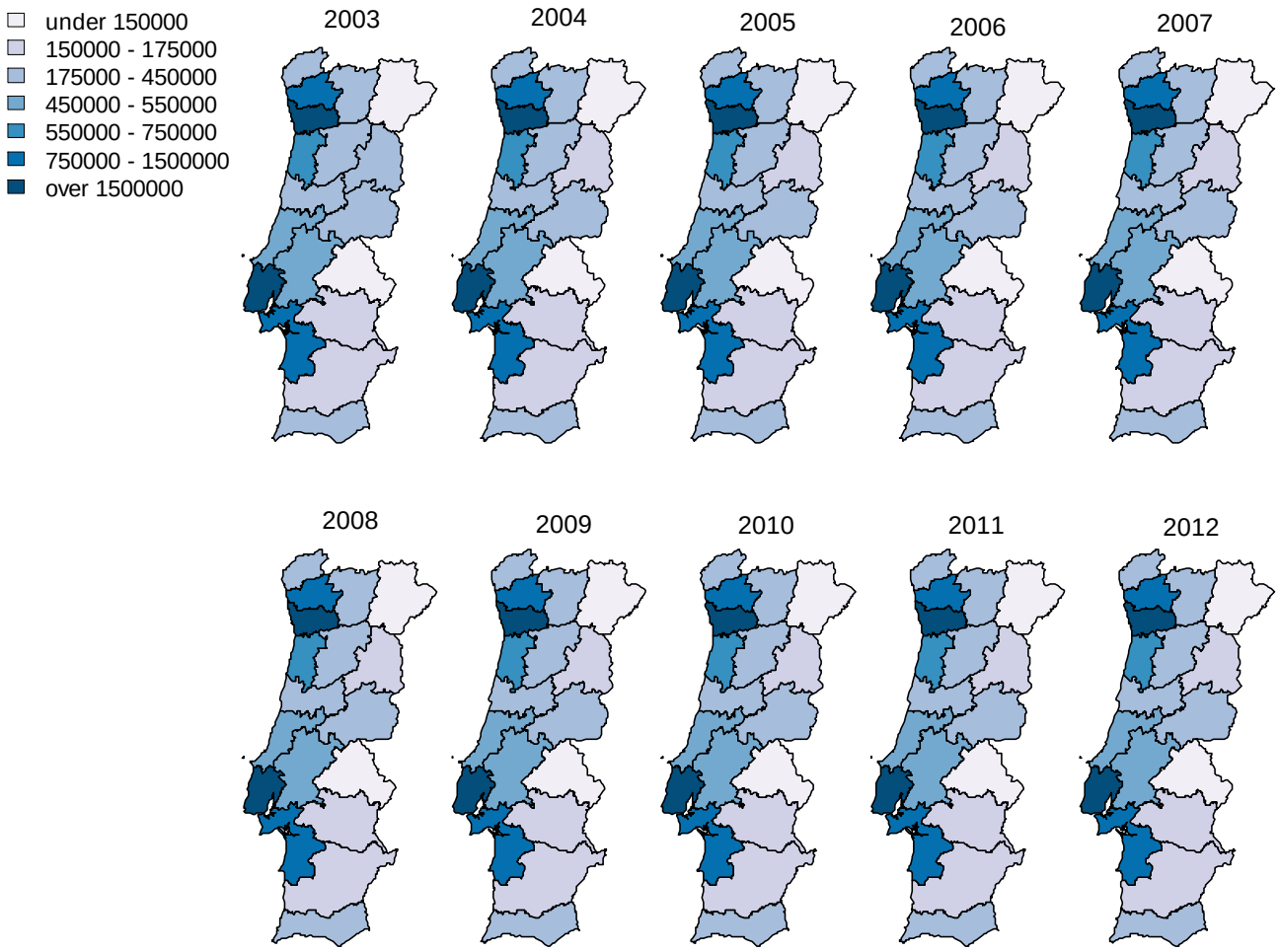


Figure 16. Evolution trend of resident population (in million inhabitants) per district of mainland Portugal.



According to the maps, the west of Portugal is more populated than the east.

There are two main populated regions: one in the north (Braga, Porto and Aveiro) and another in the south (Lisboa and Setúbal).

No differences in the population pattern were observed, over time, accordingly with the previous graphics.

Figure 17. Maps representing the resident population per district of mainland Portugal, over time.

According to previous studies, HF is the most common reason for hospital admission in people over 65 years [**9**]. The next graphics present the evolution of the proportion of population over 65 years old in Portugal. From 2003 to 2012, the proportion of people aged over 65 increased 2.5%, accounting 19.5% of the population in 2012 (Figure 18).

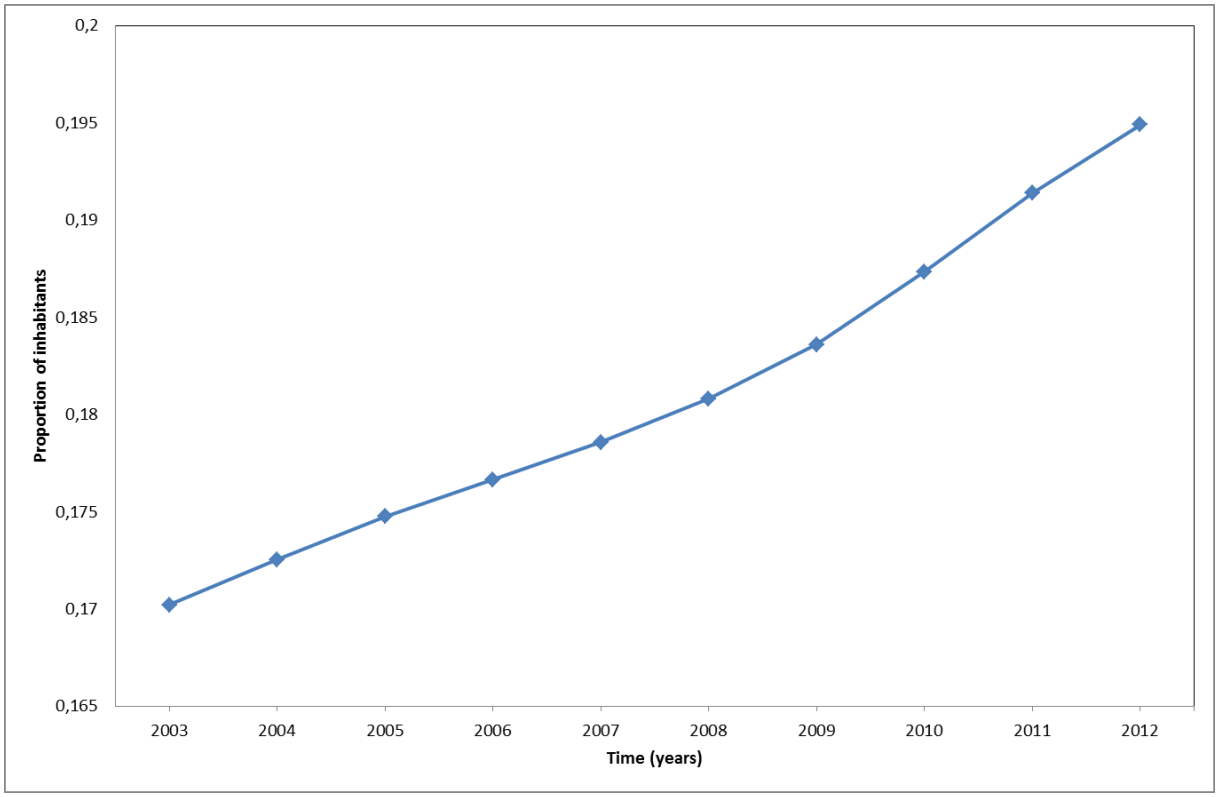


Figure 18. Proportion of population aged over 65 years old in mainland Portugal, over time.

The proportion of people aged over 65 increased between 2003 and 2012 in all the districts of mainland Portugal except in Beja and Portalegre, where it remained constant (Figure 19; Figure 20).

The most aged districts are Bragança, Guarda, Castelo Branco and Portalegre, recording also the highest number of hospital admissions per 100,000 patients.

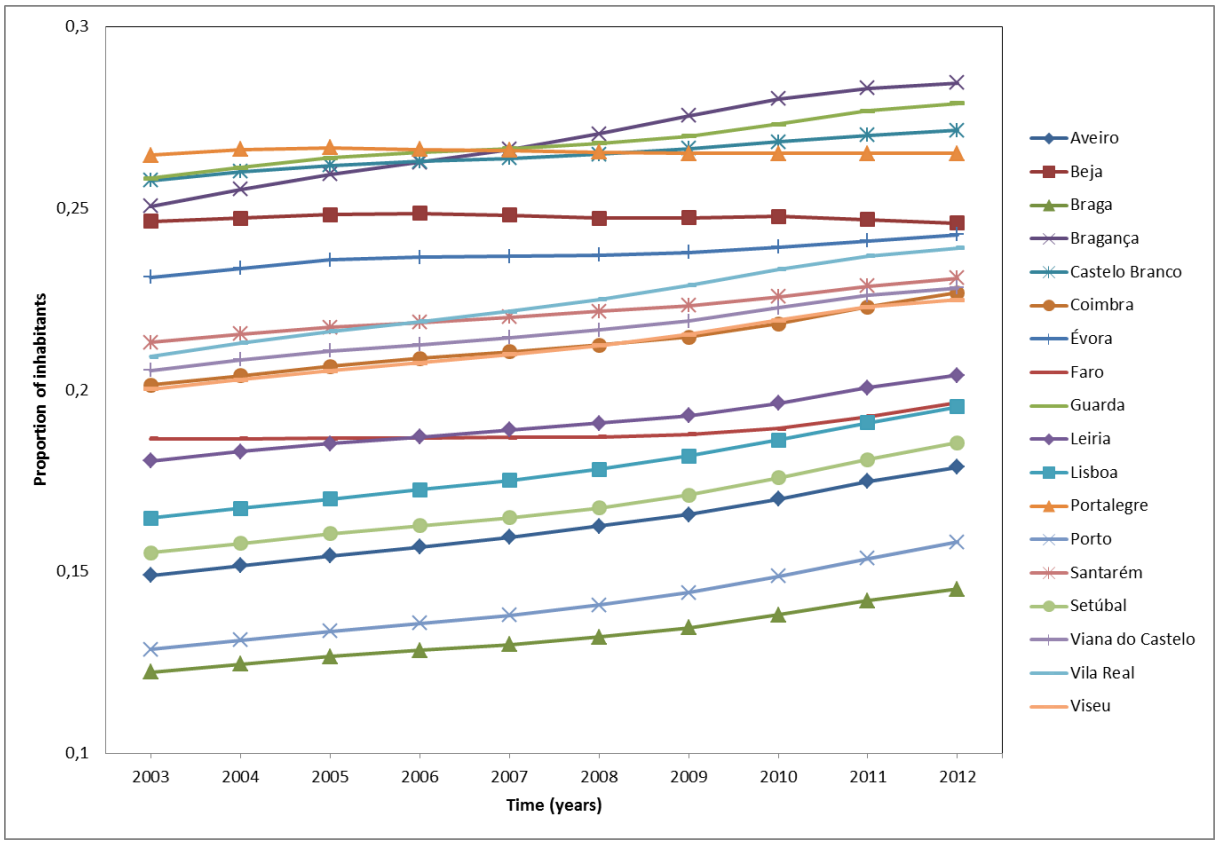


Figure 19. Proportion of population aged over 65 years old per district of mainland Portugal, over time.

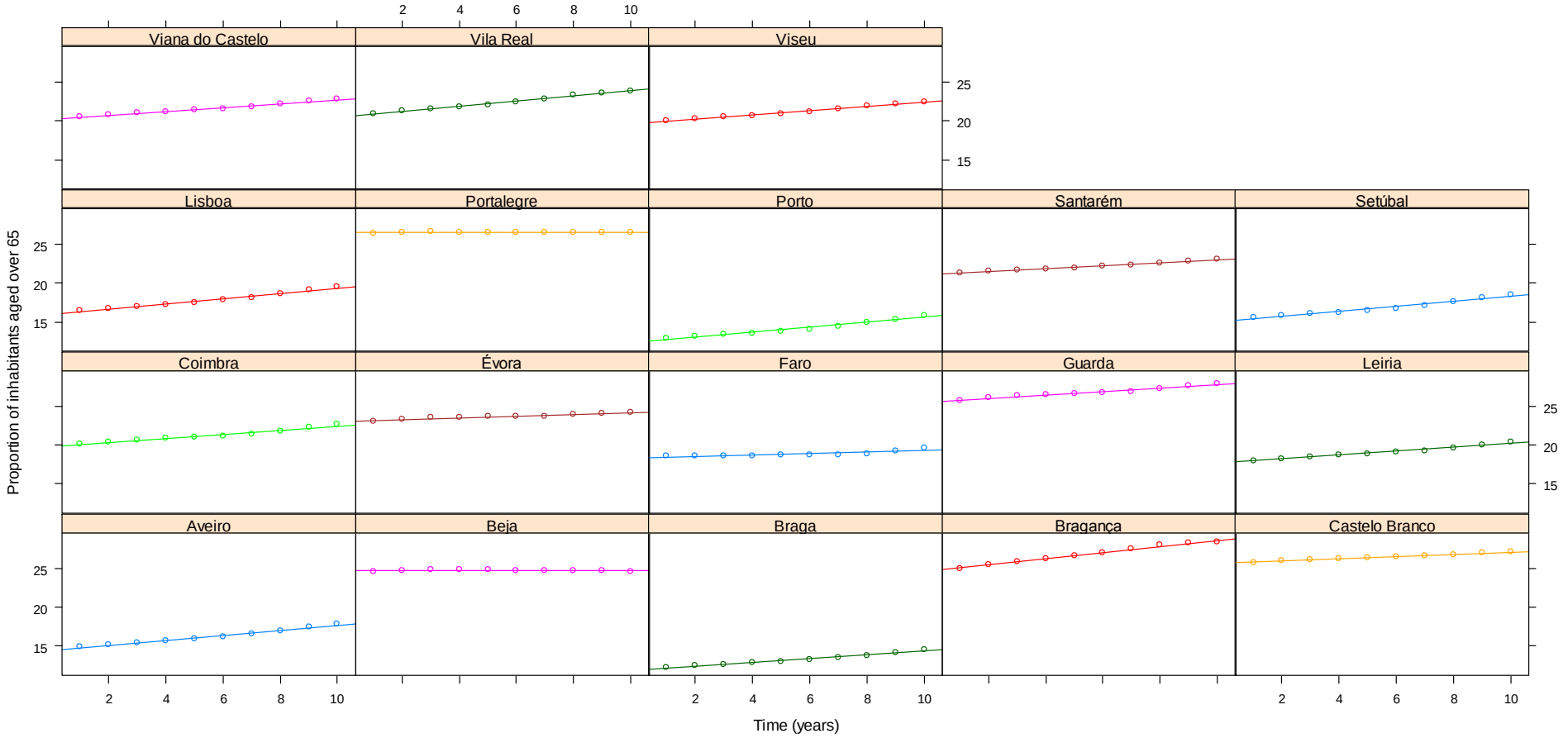
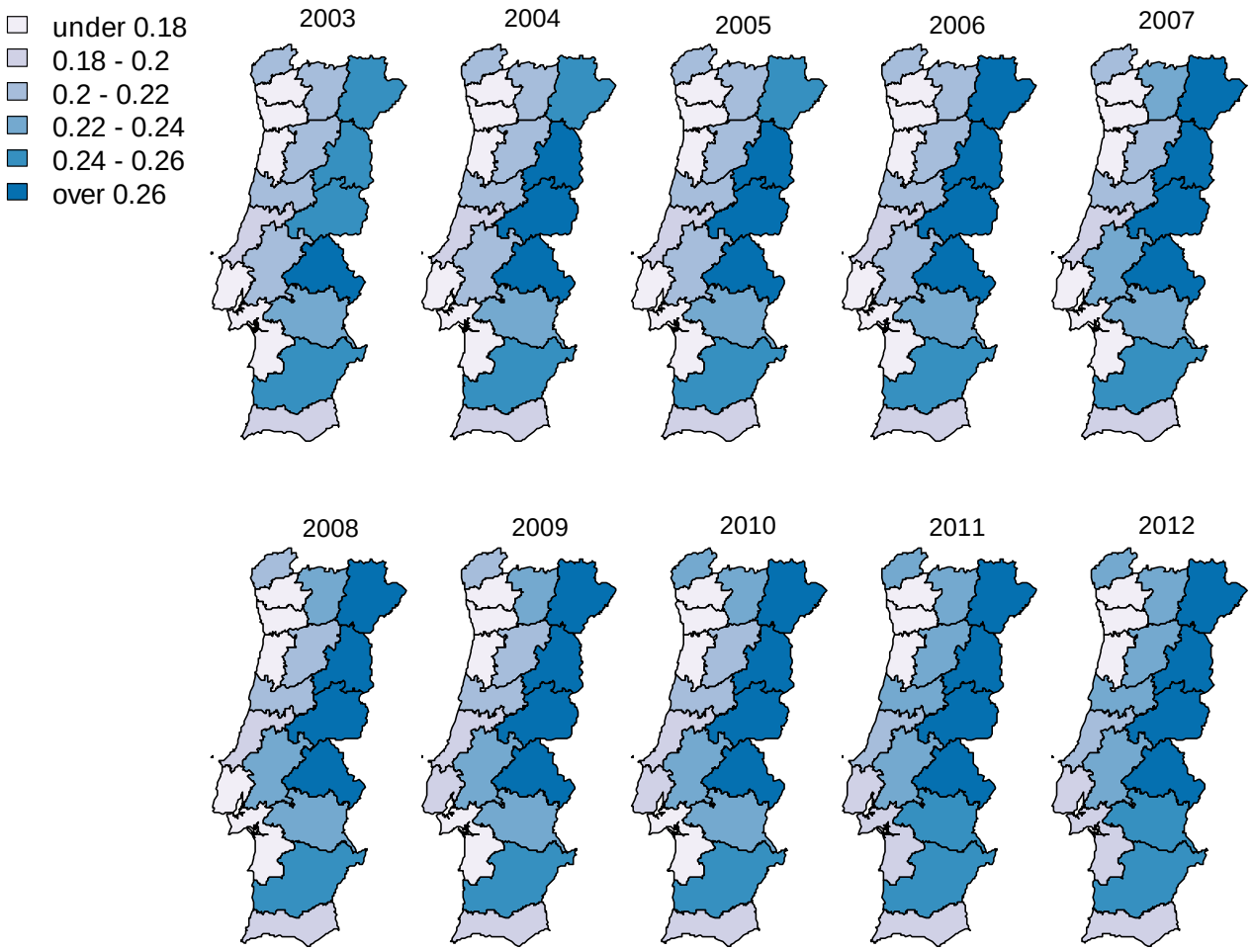


Figure 20. Evolution trend of the proportion of population aged over 65 years old per district of mainland Portugal.



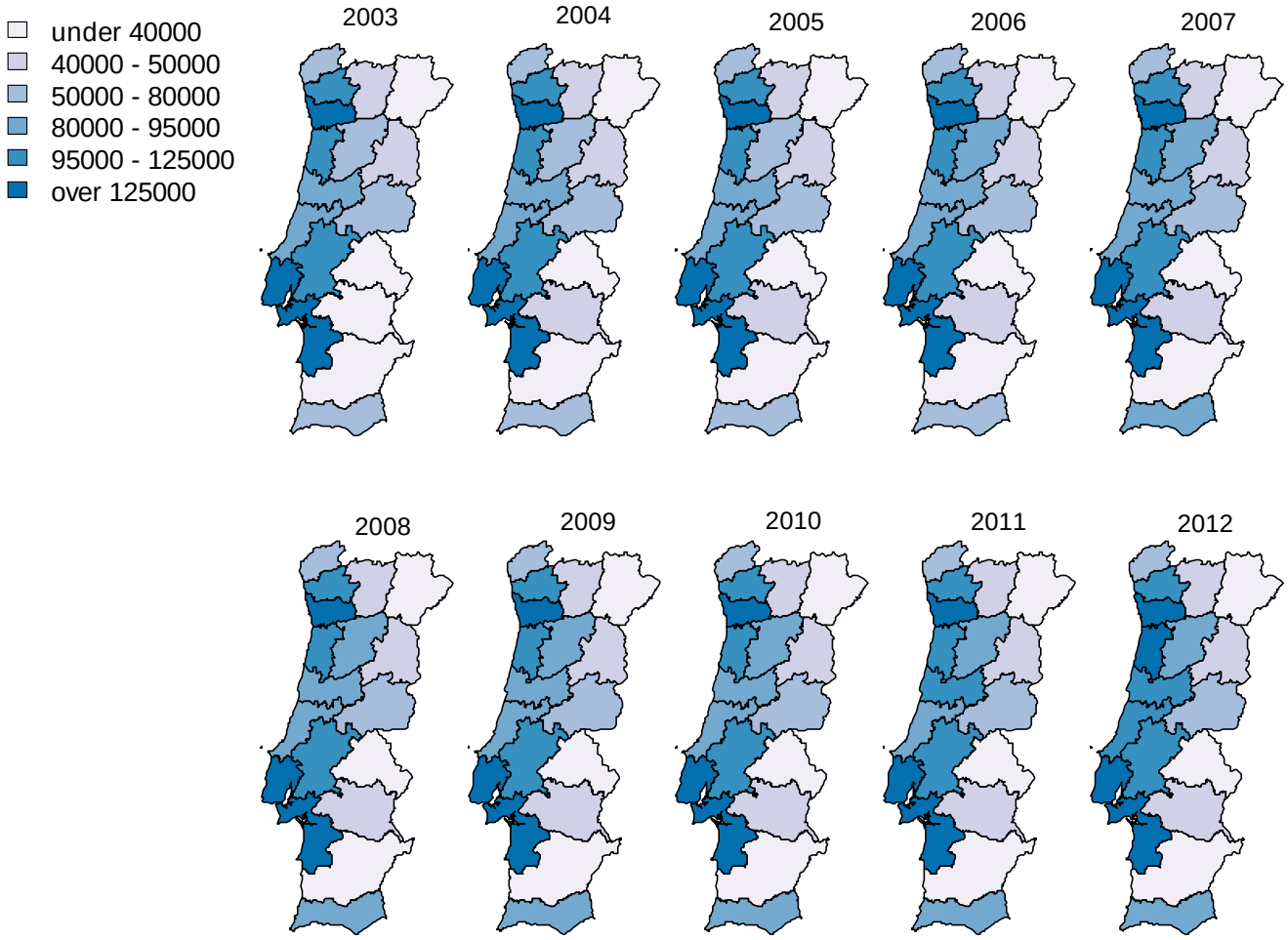
The population patterns represented in the maps support the information previously described.

The most aged districts are located in the east of mainland Portugal.

Bragança, Guarda, Castelo Branco and Portalegre are the districts with the highest proportion of population with aged over 65.

This pattern is similar to the previously presented on hospital admissions' maps (Figure 13) that showed an higher number of hospital admissions due to HF per 100,000 in the east, namely in the districts of Bragança, Guarda and Castelo Branco.

Figure 21. Maps representing the proportion of population aged over 65 years old per district of mainland Portugal, over time.



If we compare the previous maps (Figure 21) concerning the proportion of population, with the resident population aged over 65, we find an opposite pattern.

Considering the resident population, districts located in the west have higher number of aging people.

However, these districts are also the most populated, and when we compare it in relative terms, there is a higher proportion of population aged over 65 in the east of mainland Portugal.

Similarly, the districts with the highest number of hospital admissions are also located in the west of mainland Portugal (Figure 8).

Figure 22. Maps representing the resident population aged over 65 years old per district in mainland Portugal, over time.

The average proportion of population aged over 65 years increased from 17% in 2003 to 19.5% in 2012, with a 10 years average of 18.1% (Table 3).

Table 3. Average proportion of population aged over 65 years old per year.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	10 years
0,170	0,173	0,175	0,177	0,179	0,181	0,184	0,187	0,191	0,195	0,181

According to the boxplots (Figure 23) the median proportion of inhabitants aged over 65 increased over time, presenting a negative skewness in the last 5 years.

The annual average is lower than the median values and very close to the lower quartile.

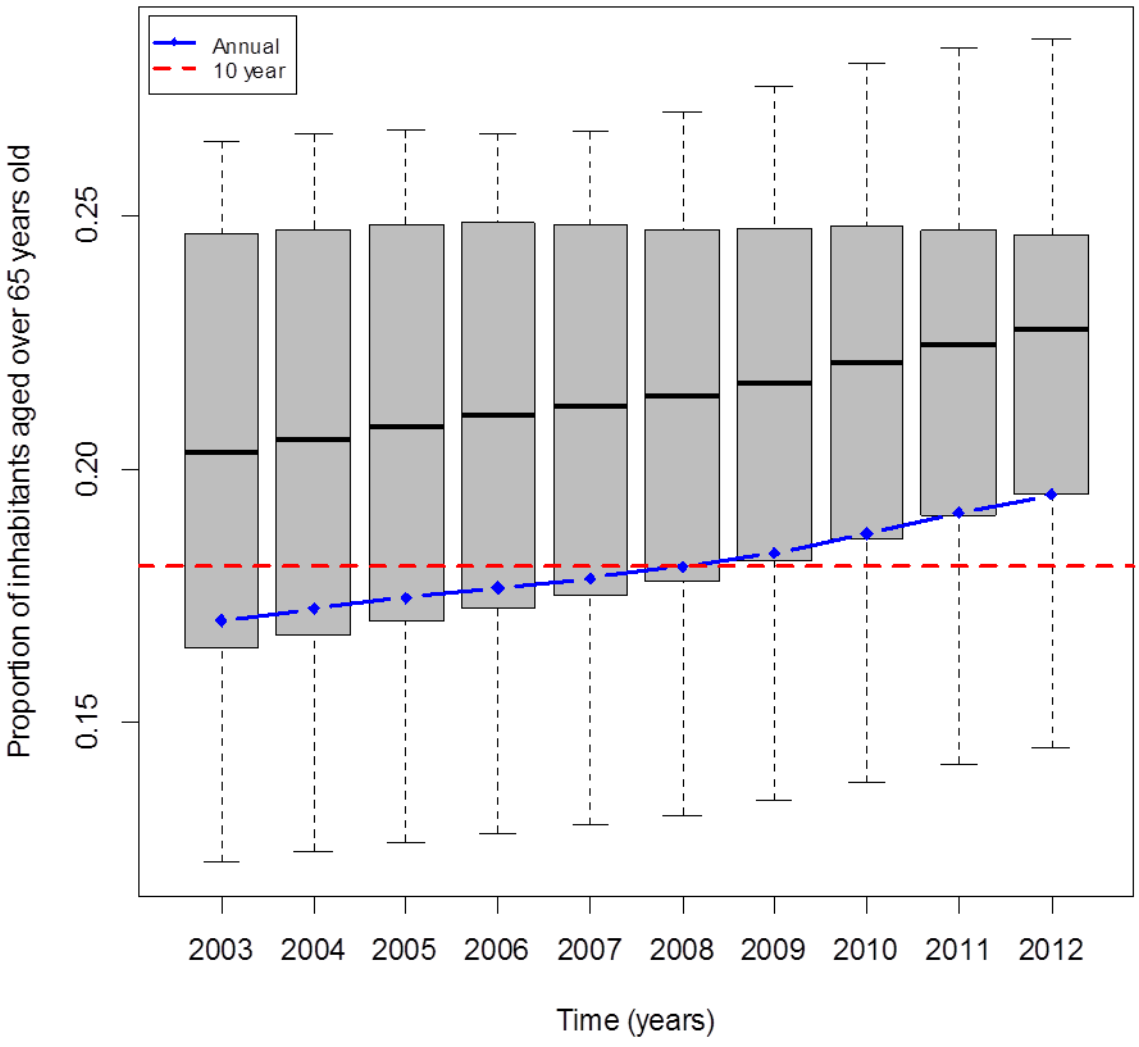


Figure 23. Boxplot with the proportion of population aged over 65 years old per district of mainland Portugal, over time.

According to the literature, elderly patients hospitalized with heart failure are mainly women [**20**]. The next graphics explore the evolution of the proportion of women in the Portuguese population aged over 65 years old.

In order to identify small fluctuations in the growing trend of the proportion of women, a reduced scale chart was used. From 2003 to 2012, the proportion of women remained approximately constant, with a variation lower than 1%, being around 58% (Figure 23).

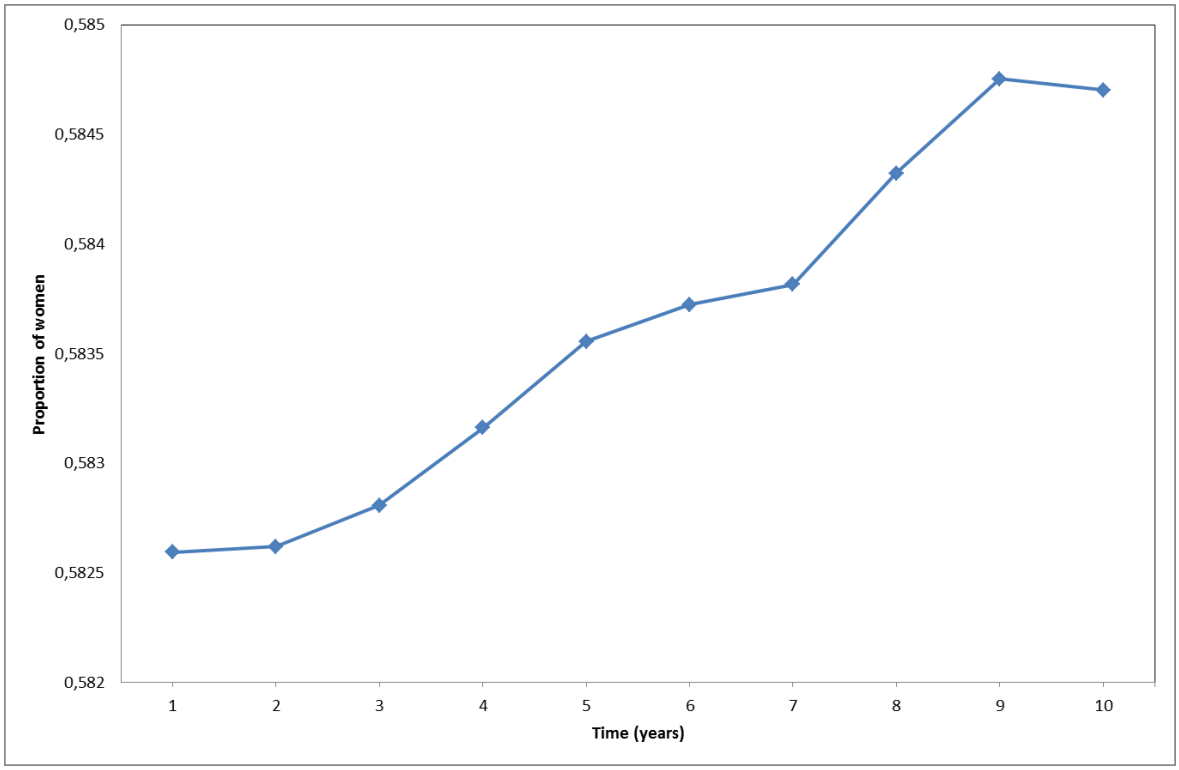


Figure 24. Proportion of women in the population aged over 65 years old in mainland Portugal, over time.

Viana do Castelo is the district with the higher proportion of women in the population aged over 65, while Faro has the lower proportion (Figure 25).

Despite 13 out of 18 countries show a growing trend, the proportion of women in the population aged over 65 remain approximately constant over time (Figure 25).

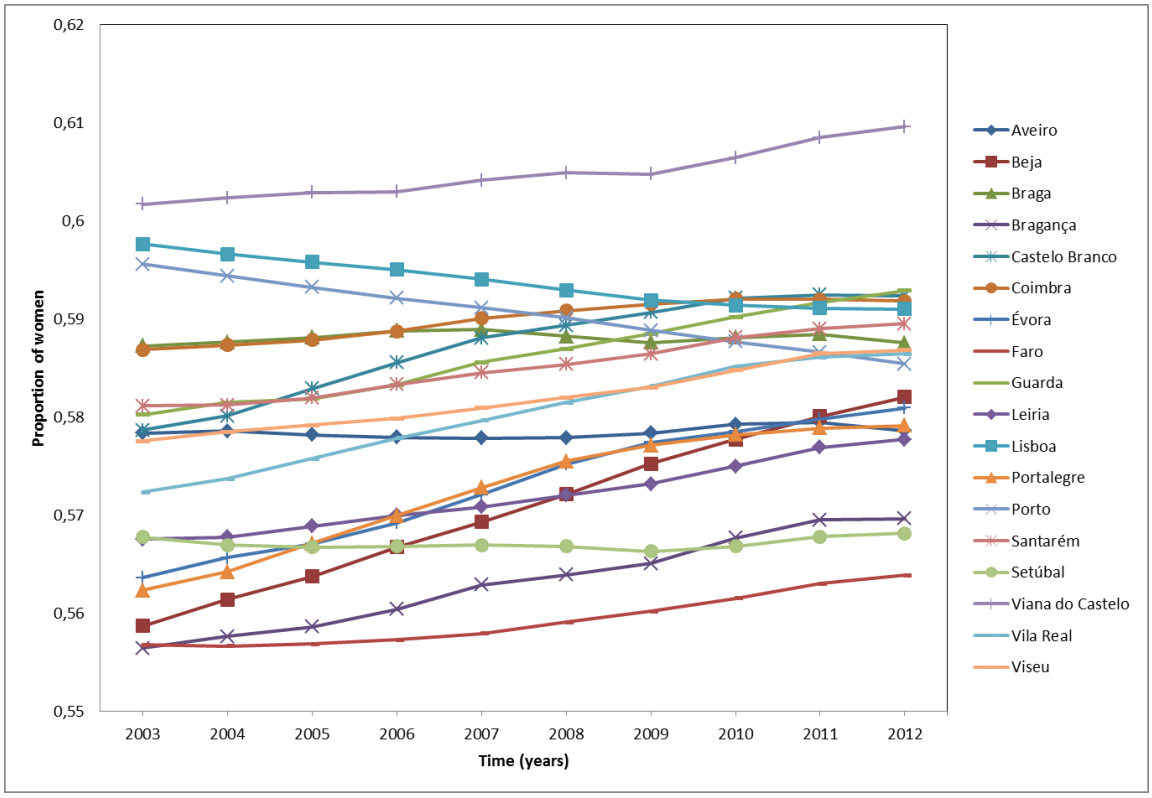


Figure 25. Proportion of women in the population aged over 65 years old per district in mainland Portugal, over time.

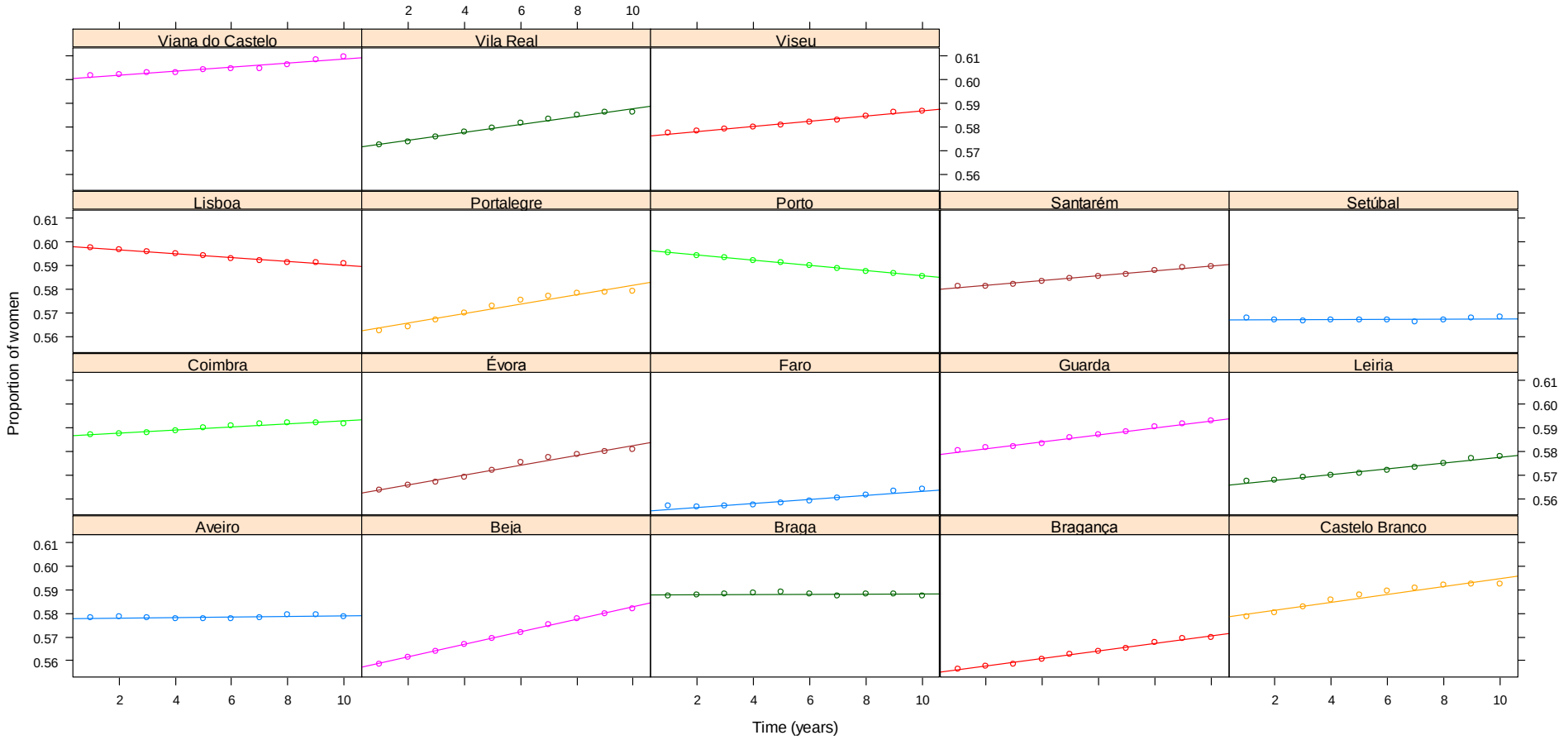
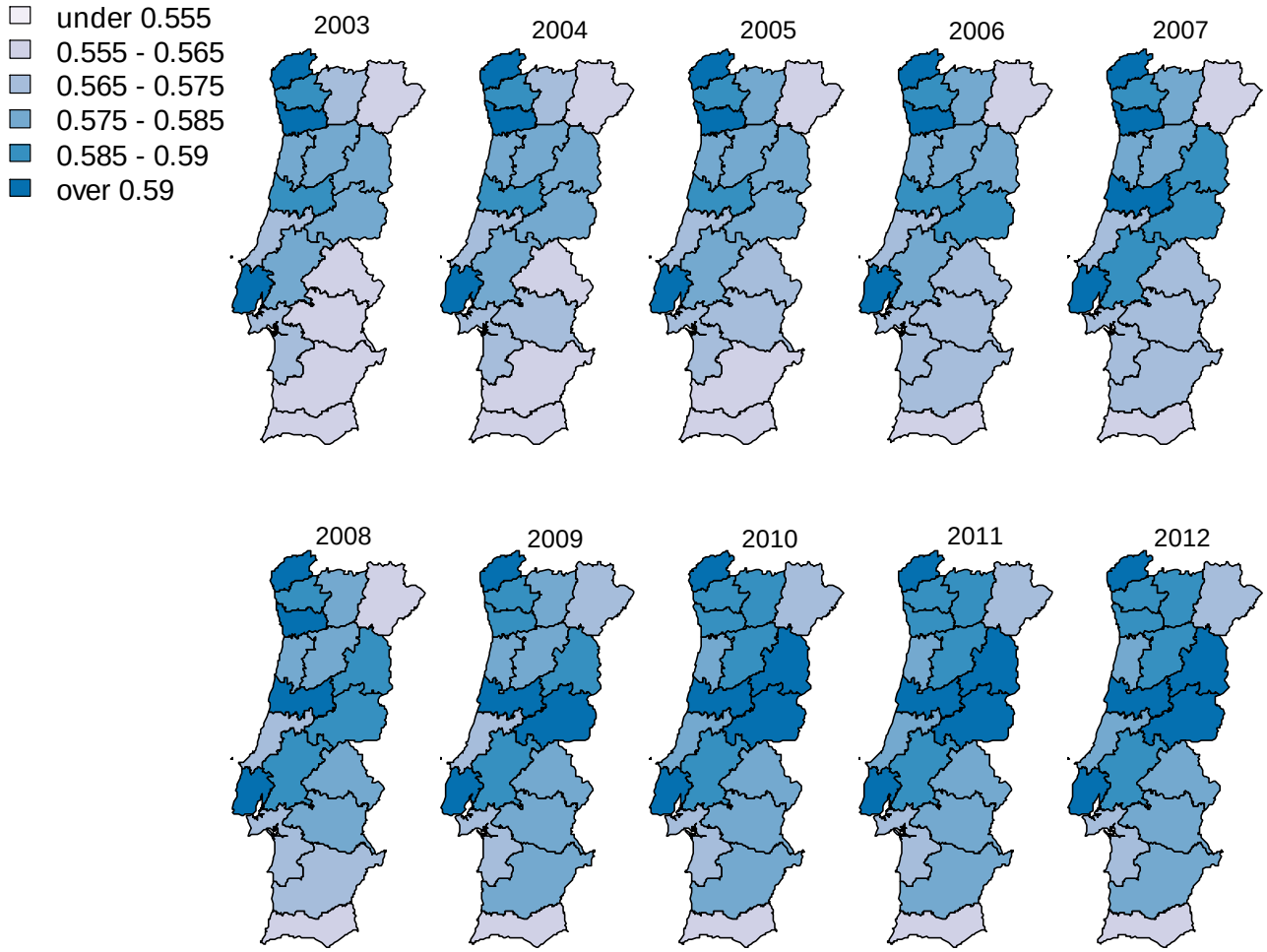


Figure 26. Evolution trend of the proportion of women in the population aged over 65 years old per district of mainland Portugal.



As expected, the proportion of women in the population aged over 65 remains constant among the various districts.

Although we used a colour scale with very small intervals, it is not possible to distinguish a clear pattern across districts, over the years.

Figure 27. Maps representing the proportion of women in the population aged over 65 years old per district in mainland Portugal, over time.

As previously described, the average proportion of women in the population aged over 65 years remains constant over the time, with an average of 58.4% (Table 4).

Table 4. Average proportion of women in the population aged over 65 years old, per year.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	10 years
0,583	0,583	0,583	0,583	0,584	0,584	0,584	0,584	0,585	0,585	0,584

According to the boxplots the average proportion of women remained stable over the years, with a small increase of the median values (Figure 28). The size of the boxplots decreases over the years suggesting that the values are converging across districts, except for Viana do Castelo.

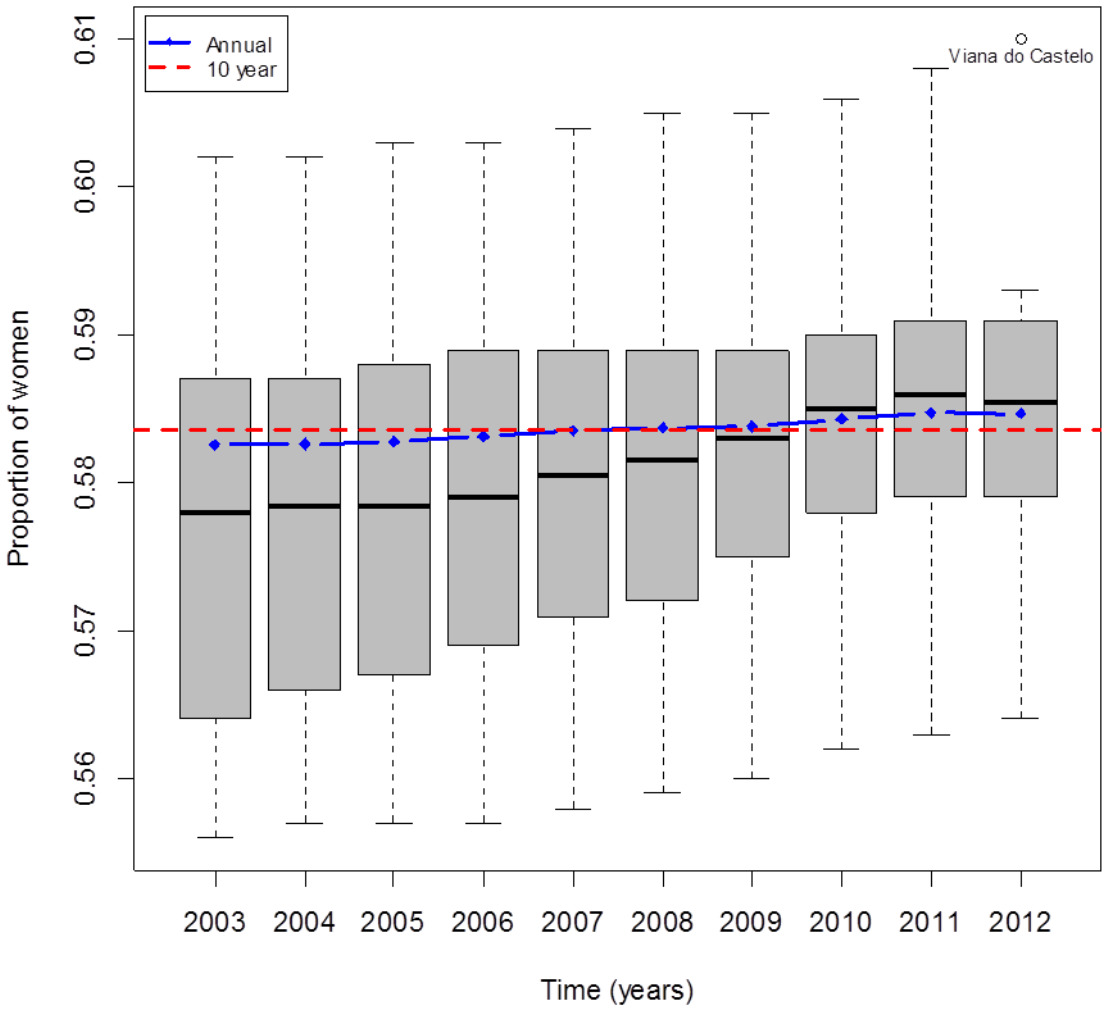


Figure 28. Boxplot with the proportion of women in the population aged over 65 years old per district of mainland Portugal, over time.

4.3. ECONOMIC FACTORS

Economic factors often influence access to healthcare. In this section we are going to study the spatial-temporal evolution of the average monthly income (AMI) in the Portuguese population.

Between 2003 and 2012 there was an increase of 28% in the AMI of the Portuguese population (Figure 29).

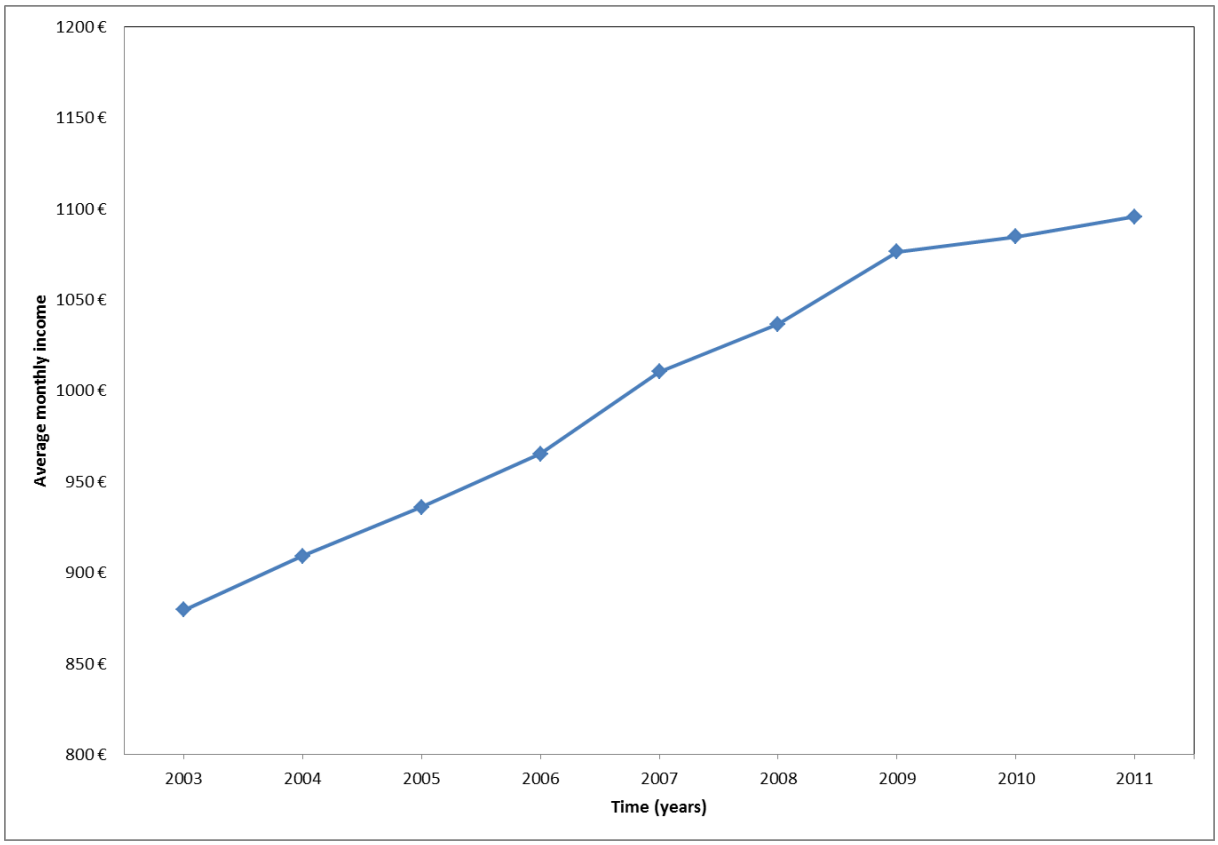


Figure 29. Average monthly income of the Portuguese population, over time.

Lisboa and Setúbal are the districts with the highest increase, reaching in 2012 an AMI of 1,308€ and 1,111€, respectively (Figure 30).

Nevertheless, all the 18 Portuguese districts showed a constant increase in the AMI, over the 10 years (Figure 31).

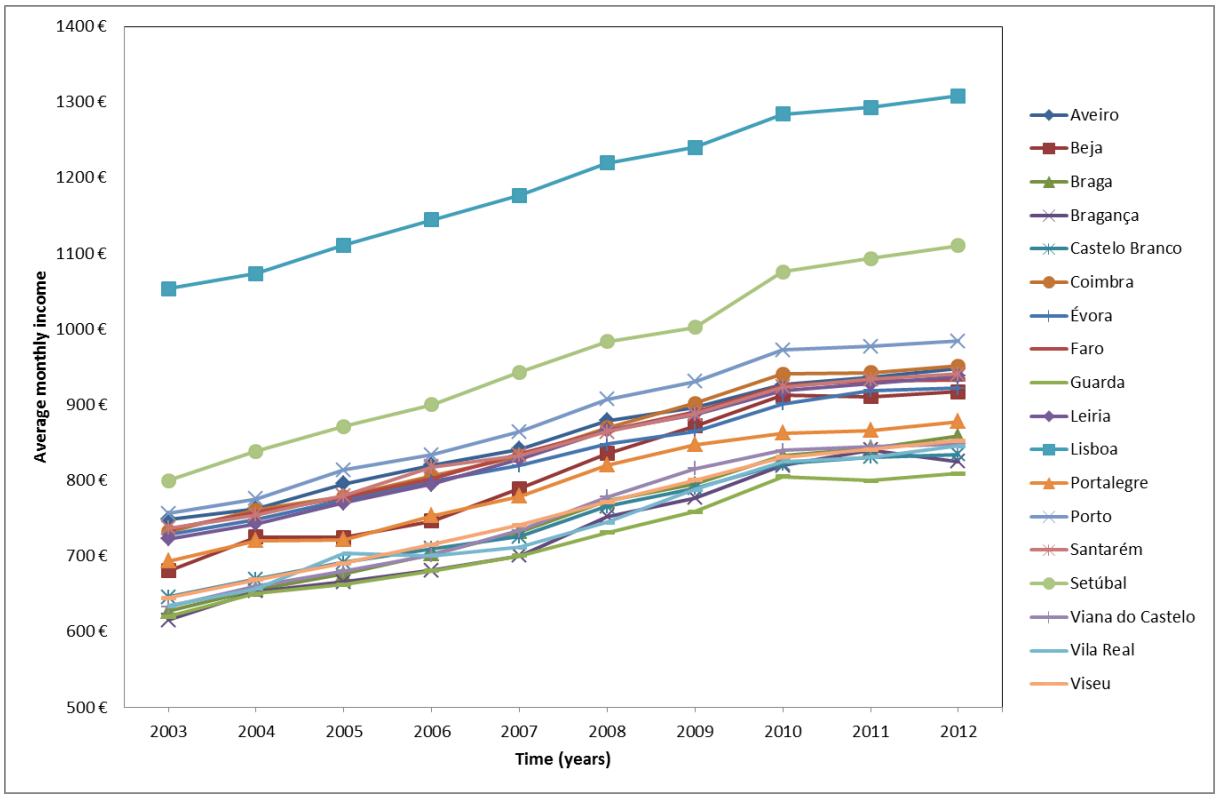


Figure 30. Average monthly income of the Portuguese population per district, over time.

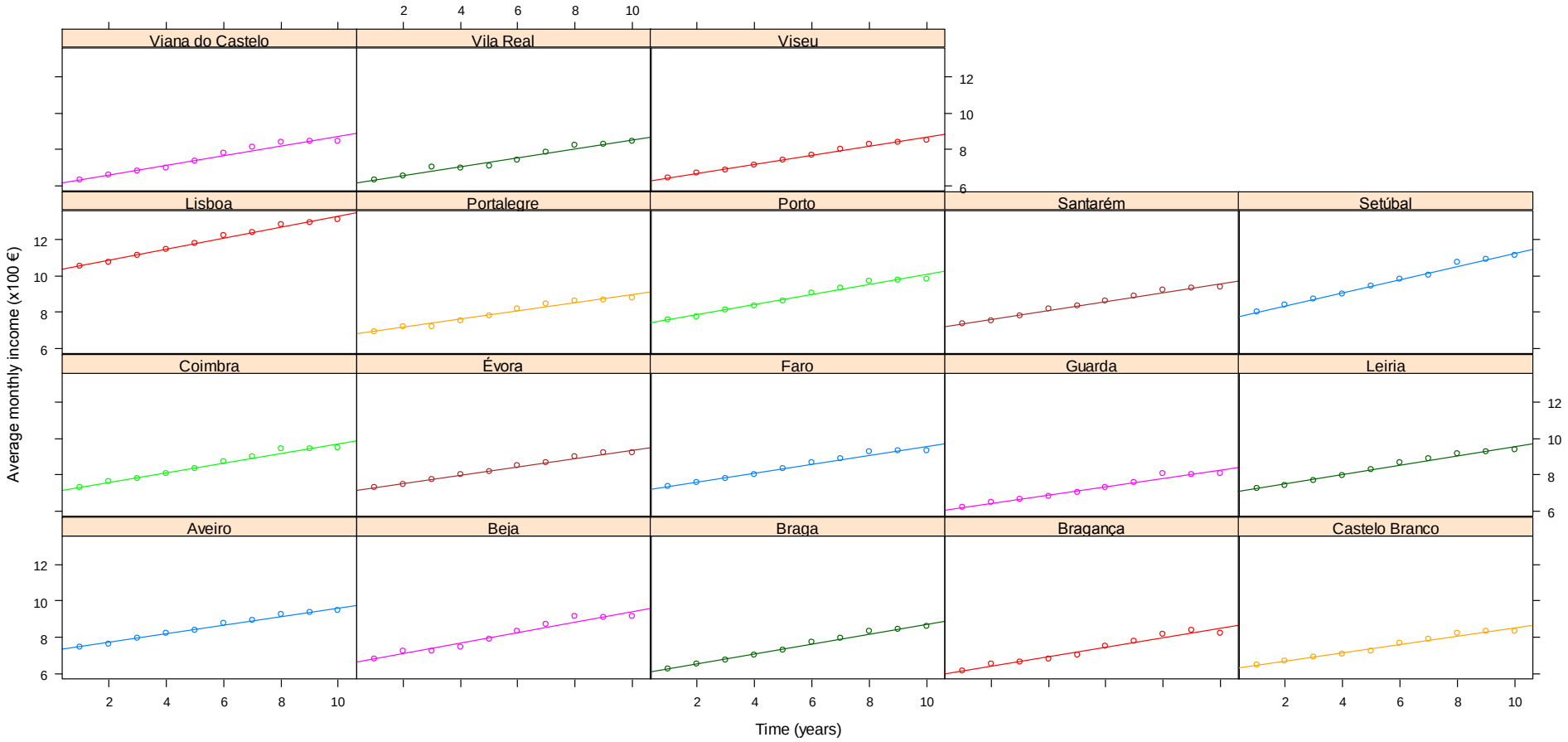


Figure 31. Evolution trend of the average monthly income, per district of mainland Portugal.

The AMI of the Portuguese population increased from 854.1€ in 2003 to 1,095,6€ in 2012, with a 10 years average of 984,7 (Table 5).

Table 5. Average monthly income in mainland Portugal, over time.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	10 years
854,1	879,6	909,2	936,0	965,3	1010,4	1036,4	1076,3	1084,6	1095,6	984,7

In the boxplot (Figure 32) it is possible to identify two outliers that positively influence the AMI: Lisboa (over the 10 years) and Setúbal (between 2011 and 2012), which supports the results previously described.

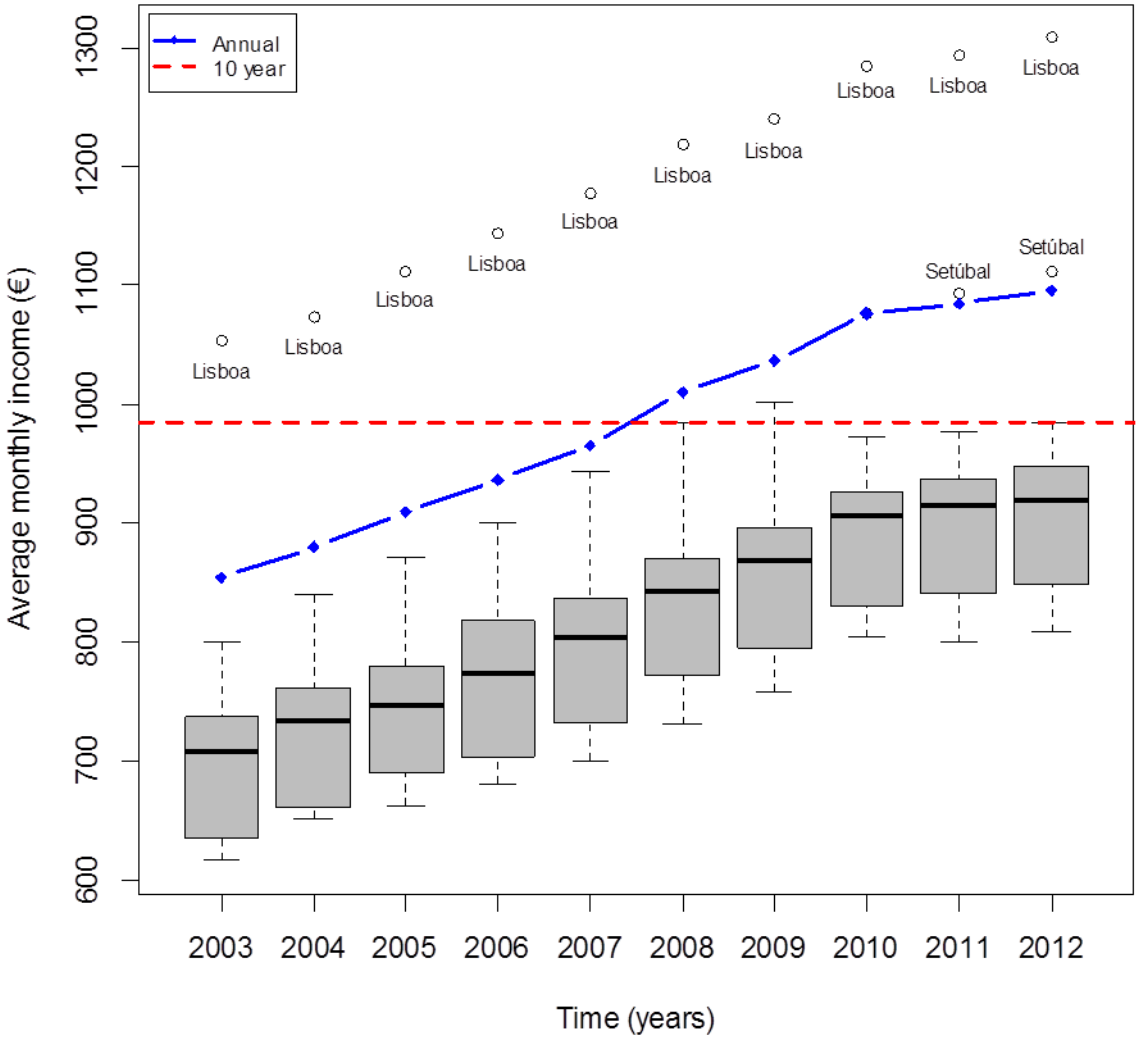


Figure 32. Boxplot with the average monthly income, per district of mainland Portugal.

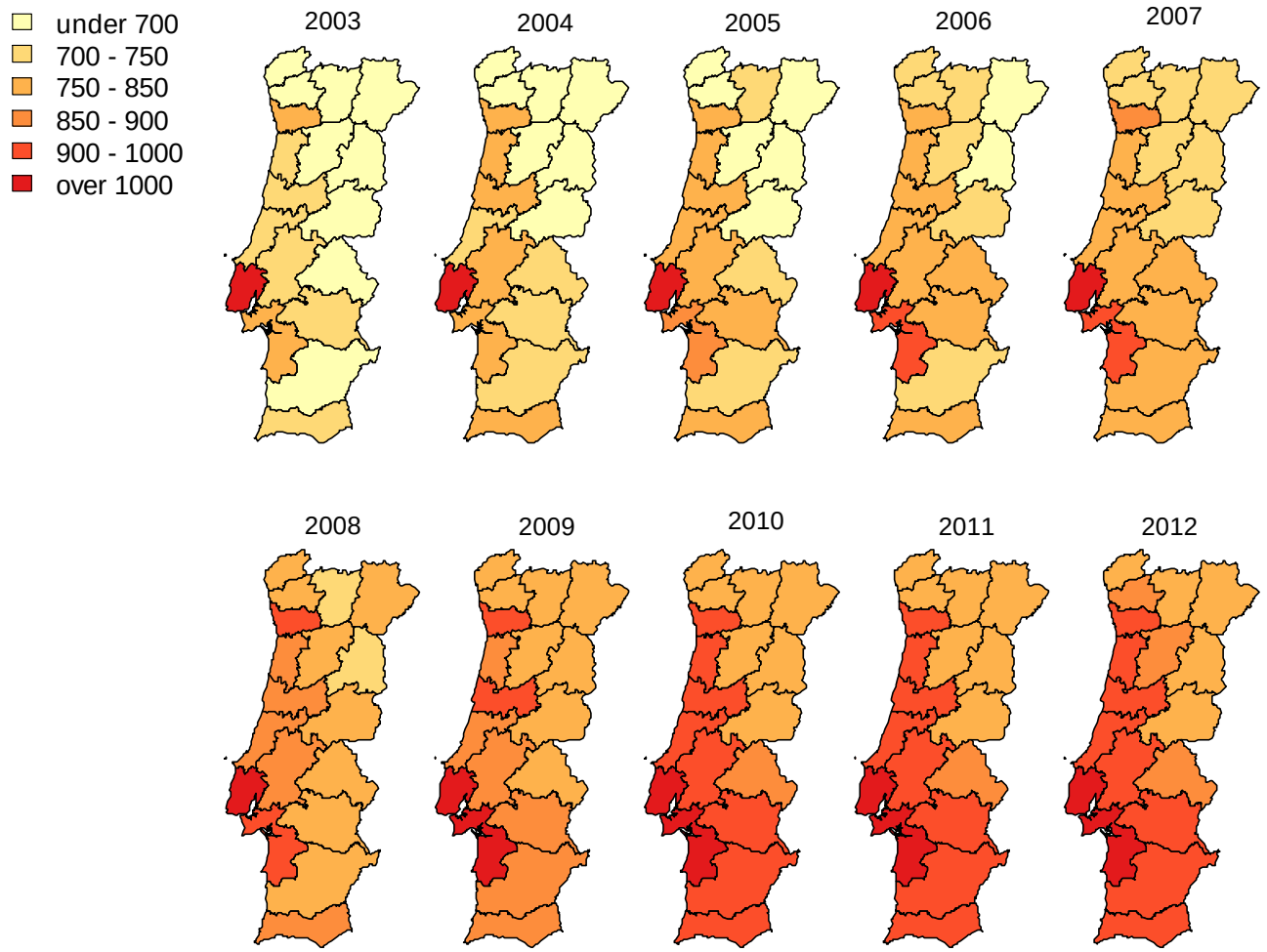


Figure 33. Maps representing the average monthly income (in euros) of the Portuguese population per district, ove

As previously described, the district with the highest AMI is Lisboa, followed by Setúbal.

However, from 2010, the districts of west and south show a similar pattern, with a higher AMI compared to the others.

This pattern is opposite to the distribution of the number of hospital admissions per 100,000 inhabitants (Figure 8).

4.4. SOCIAL FACTORS

The number of hospitals and primary care centres per 100,000 inhabitants allows measuring the access of population to healthcare services.

The next graphics explore the evolution of hospitals and primary care centres in Portugal, between 2003 and 2012.

The number of hospitals per 100,000 inhabitants remained constant over the years, with minor fluctuations identified by the reduced scale used in the chart (Figure 34).

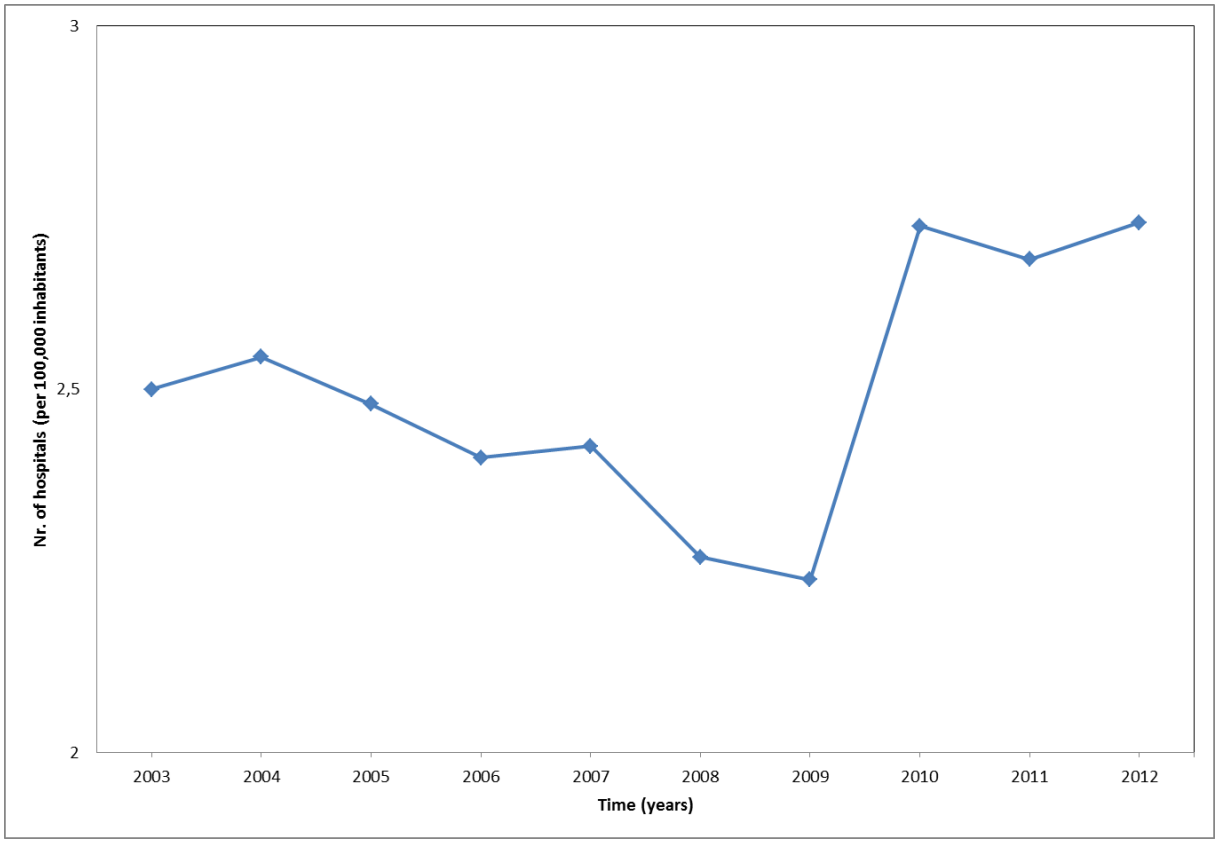


Figure 34. Number of hospitals per 100,000 inhabitants in mainland Portugal, over time.

Lisboa and Coimbra are the districts with the higher number of hospitals per 100,000 inhabitants, reaching in 2012 5.2 and 4.5 respectively (Figure 35).

In the other districts the number of hospitals remained approximately constant over the years (Figure 36).

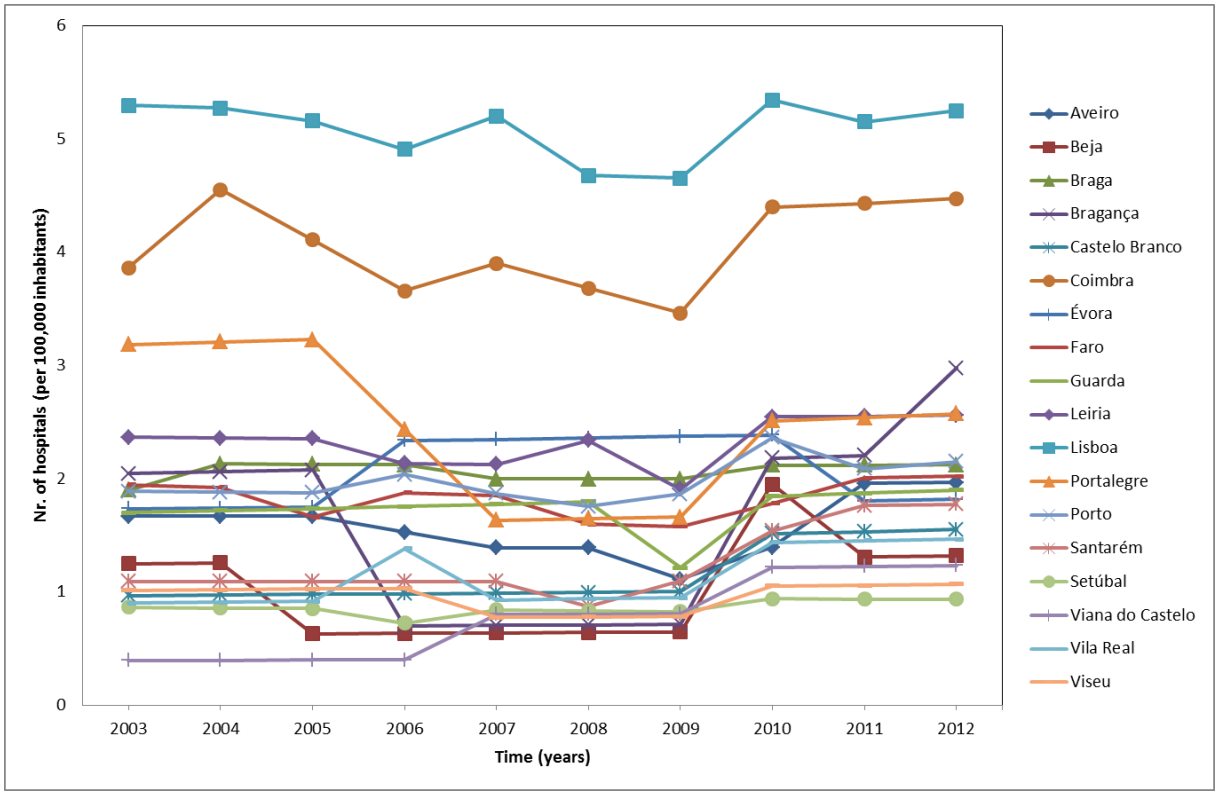


Figure 35. Number of hospitals per 100,000 inhabitants per district in mainland Portugal, over time.

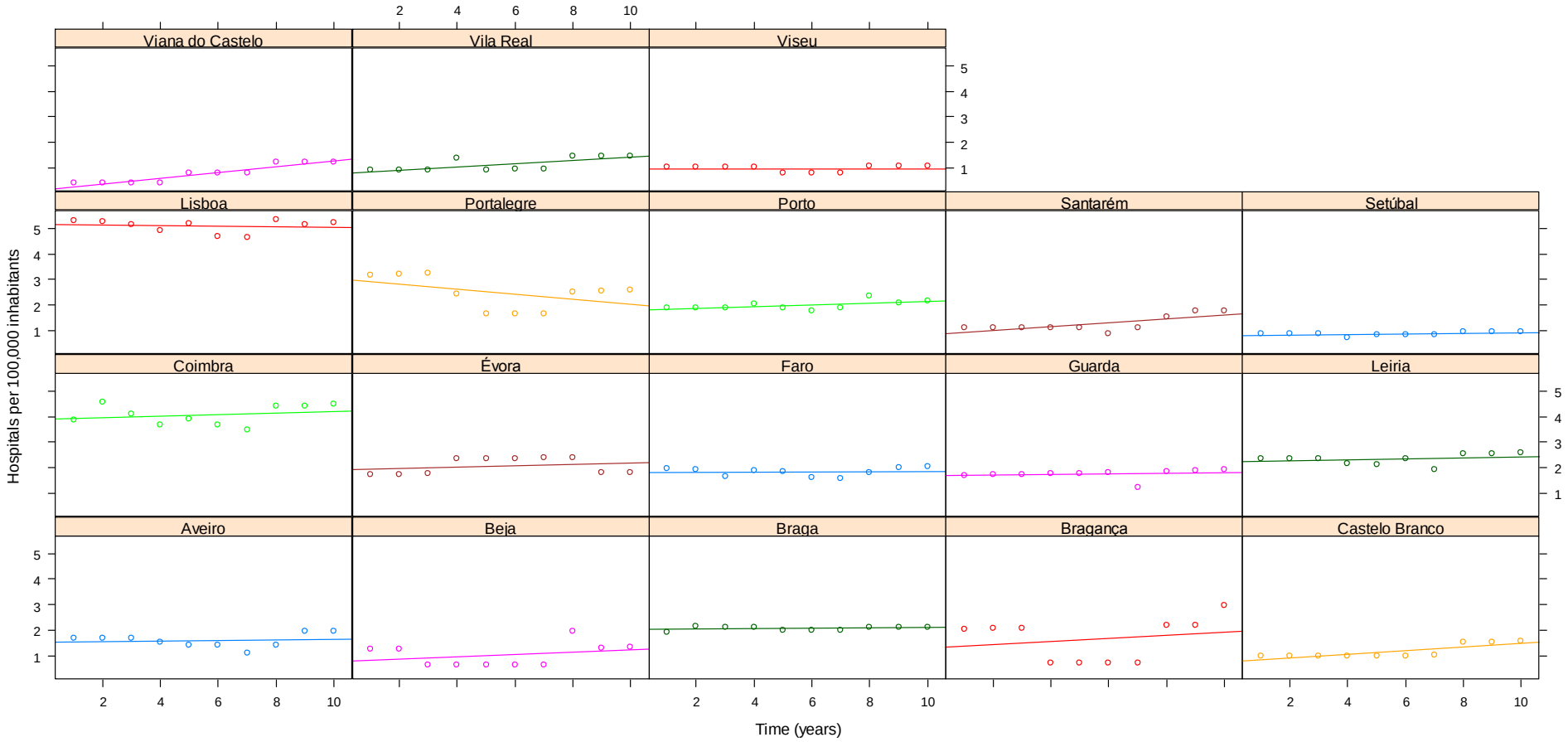


Figure 36. Evolution trend of hospitals per 100,000 inhabitants per district in mainland Portugal.

The average number of hospital per 100,000 inhabitants over the 10 years was 2.5 (Table 6).

Table 6. Average number of hospital per 100,000 inhabitants in mainland Portugal, over time.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	10 years
2,5	2,5	2,5	2,4	2,4	2,3	2,2	2,7	2,7	2,7	2,5

In the boxplot (Figure 37) it is possible to identify two outliers, Lisboa and Coimbra that positively influence the average number of hospital per 100,000 inhabitants, which supports the results previously described.

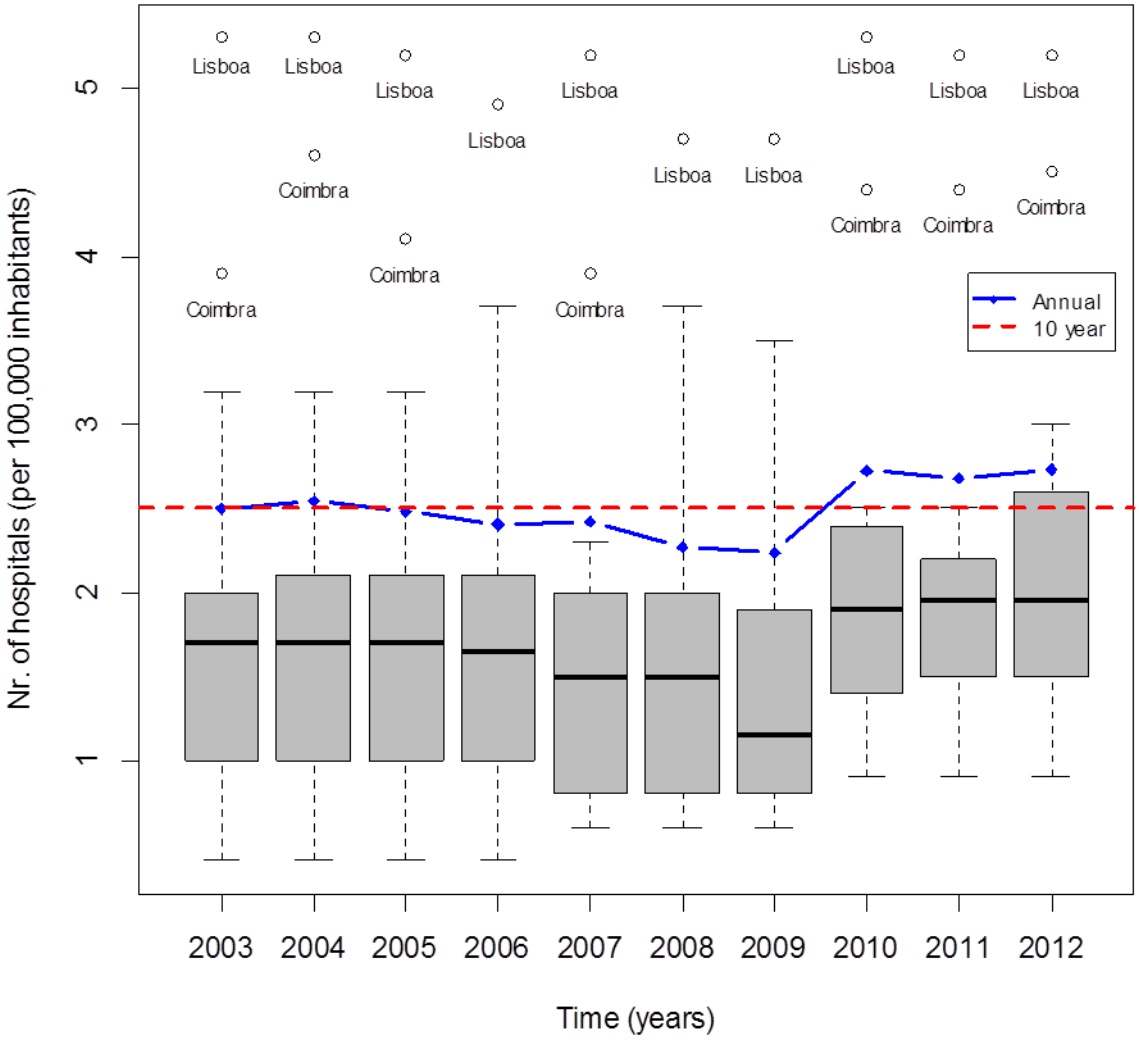
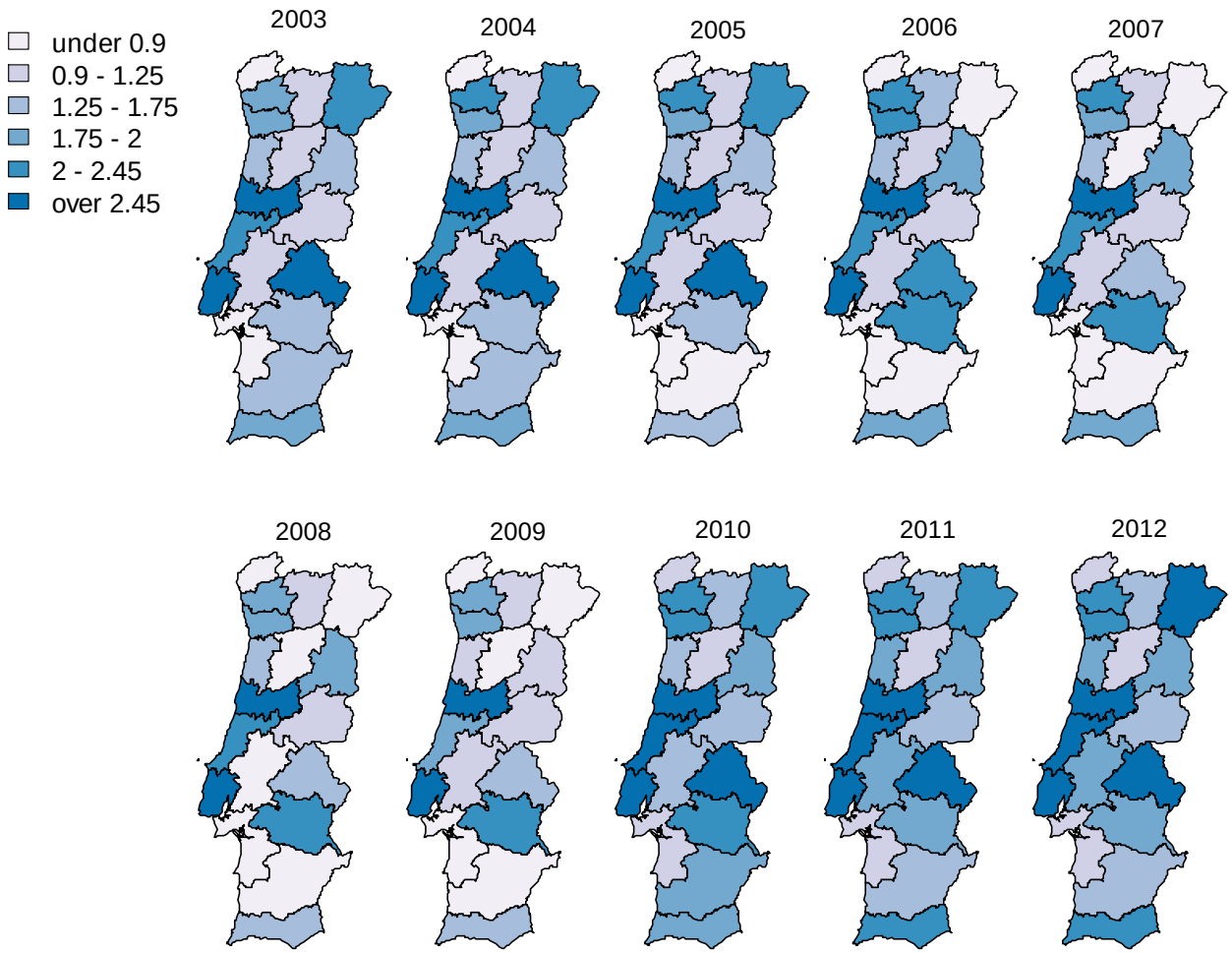


Figure 37. Boxplot with the number of hospital per 100,000 inhabitants in mainland Portugal, over time.



The maps distribution pattern is consistent with the previous results.

Districts of west have a higher number of hospitals per 100,000 inhabitants.

Figure 38. Maps representing the number of hospitals per 100,000 inhabitants per district in mainland Portugal, c

Primary care centres play a determinant role in the diagnosis and management of patients with HF.

Between 2003 and 2012, the number of primary care centres per 100,000 inhabitants remained constant (Figure 39).

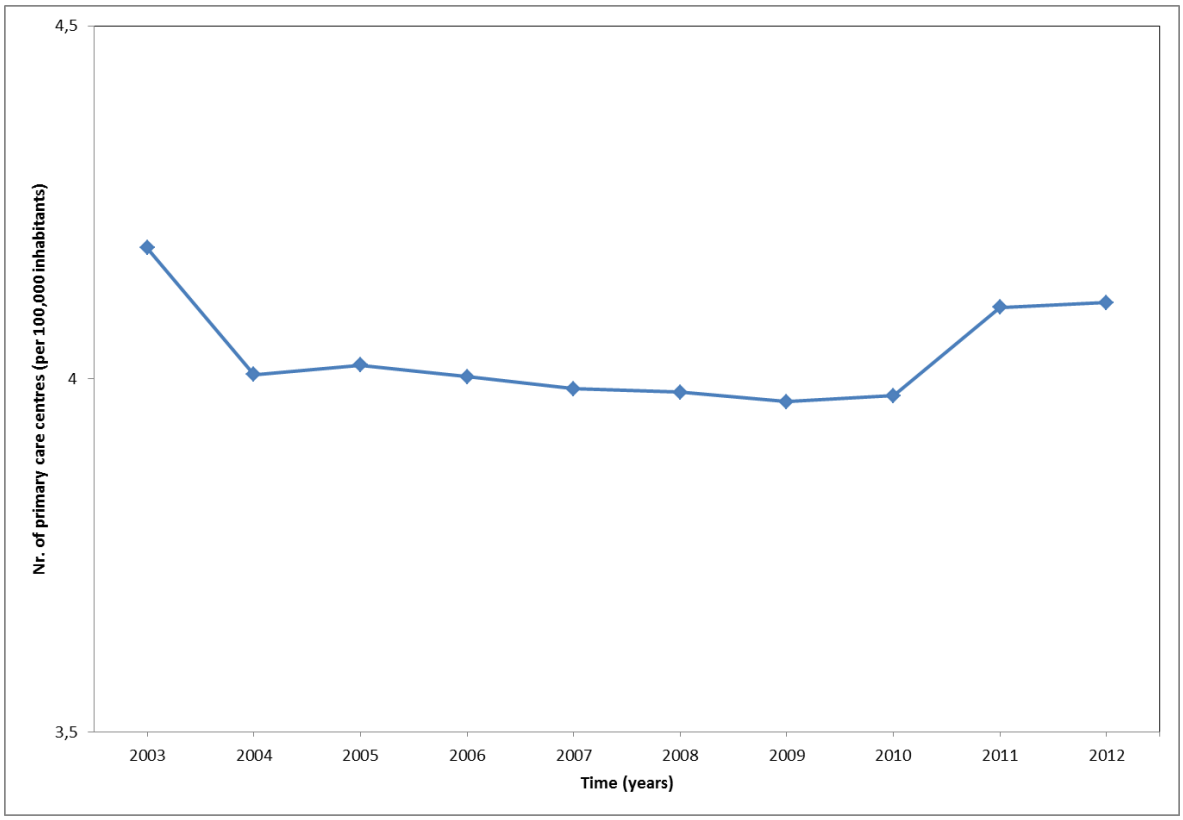


Figure 39. Number of primary care centres per 100,000 inhabitants in mainland Portugal, over time.

Portalegre is the only district with more than 12 primary care centres per 100,000 inhabitants, followed by Beja, Bragança, Guarda and Évora (Figure 40).

In the opposite, Braga, Porto, Setúbal and Aveiro are the districts with the lower number, below 4 per 100,000 inhabitants.

In all the districts, the number of primary care centres approximately constant over the years (Figure 41).

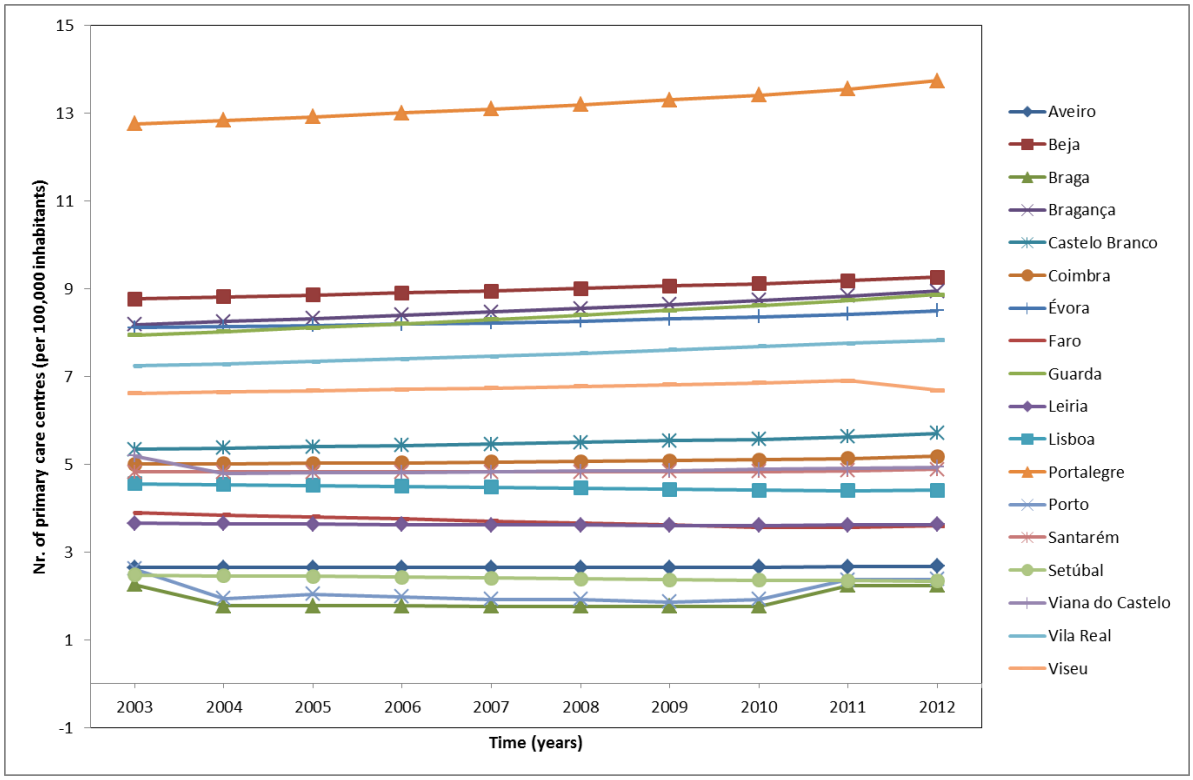


Figure 40. Number of primary care centres per 100,000 inhabitants per district in mainland Portugal, over time.

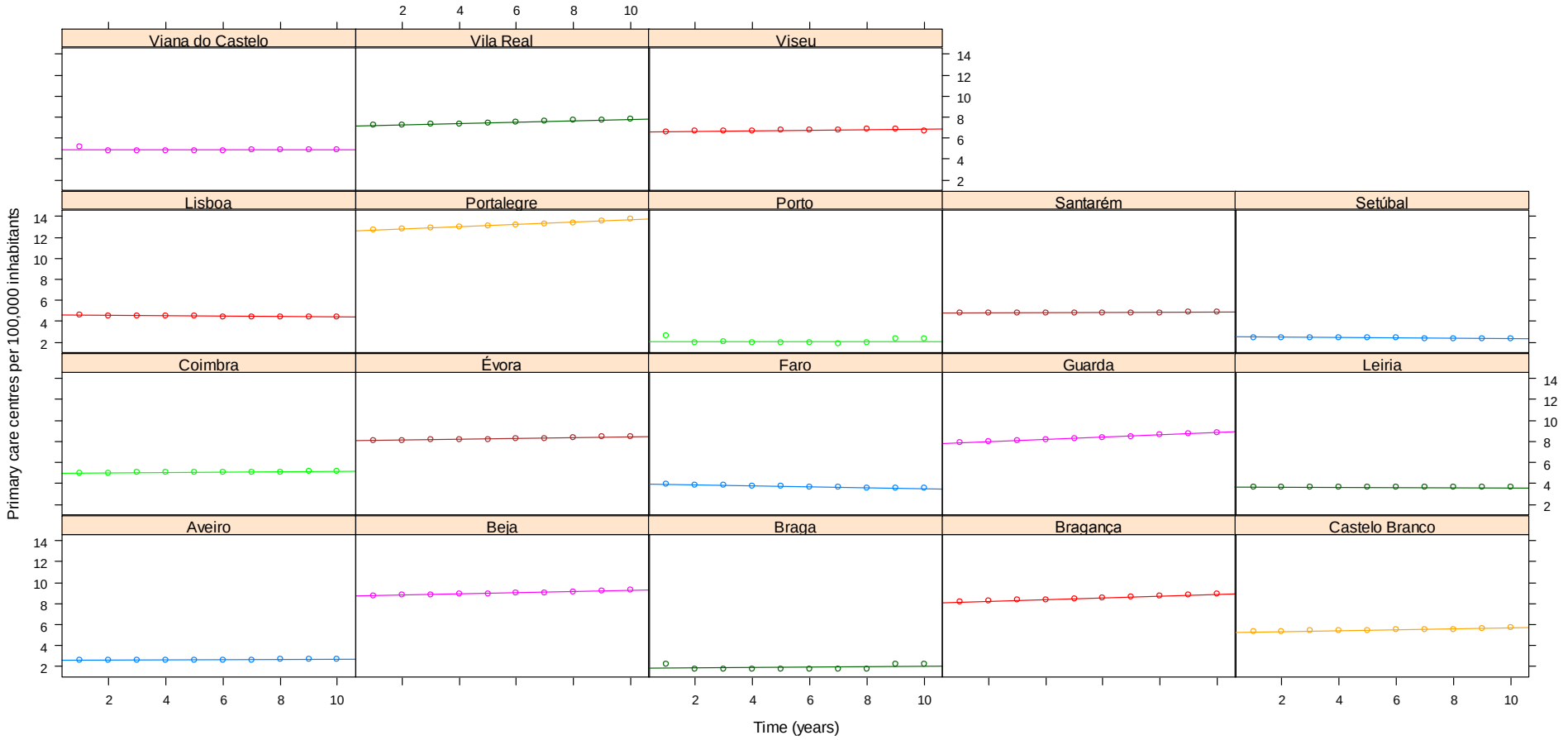


Figure 41. Evolution trend of primary care centres per 100,000 inhabitants per district of mainland Portugal.

The average number of primary care centres per 100,000 inhabitants was 4 (Table 6).

Table 7. Average number of primary care centres per 100,000 inhabitants in mainland Portugal, over time.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	10 years
4,2	4,0	4,0	4,0	4,0	4,0	4,0	4,0	4,1	4,1	4,0

The annual average and the median number of primary care centres remained stable over the years (Figure 42).

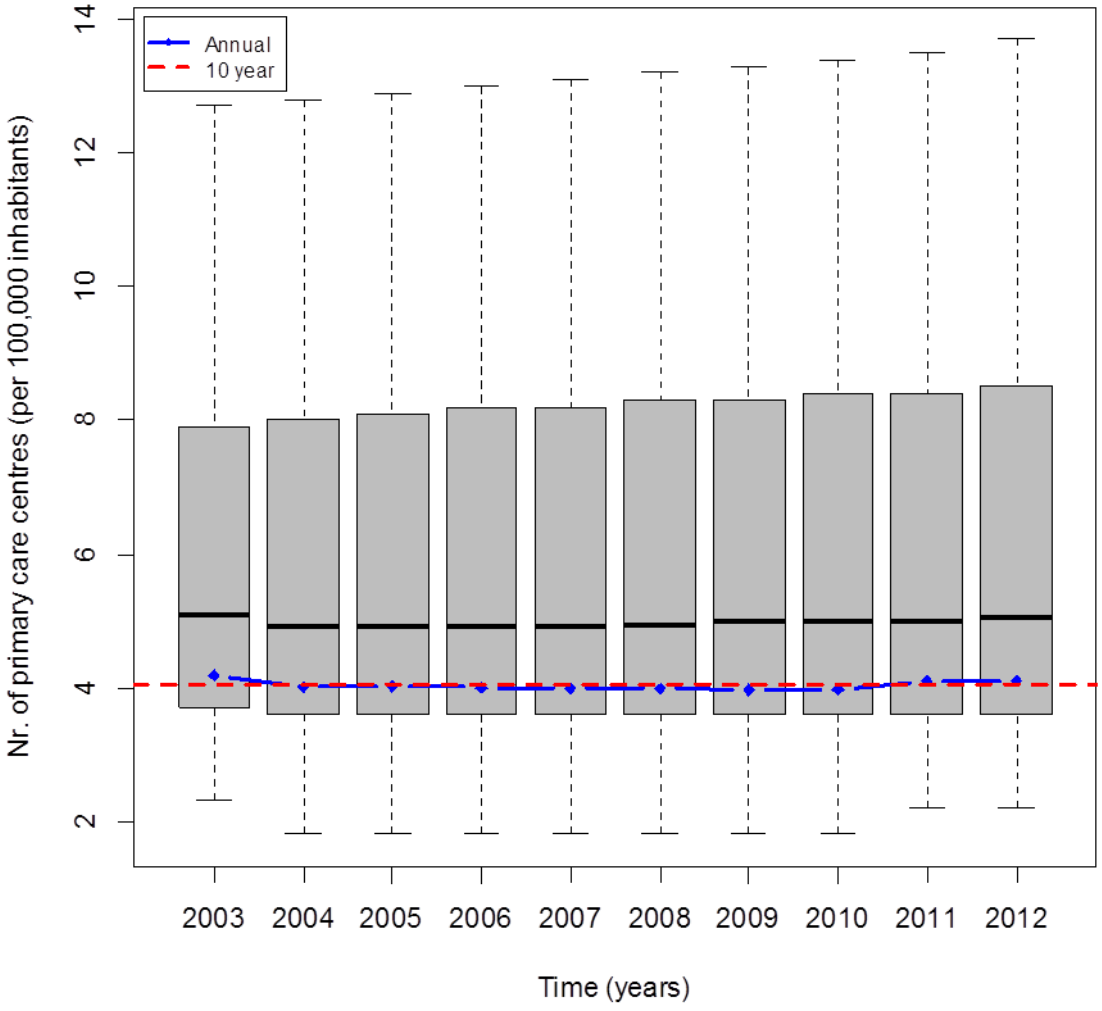
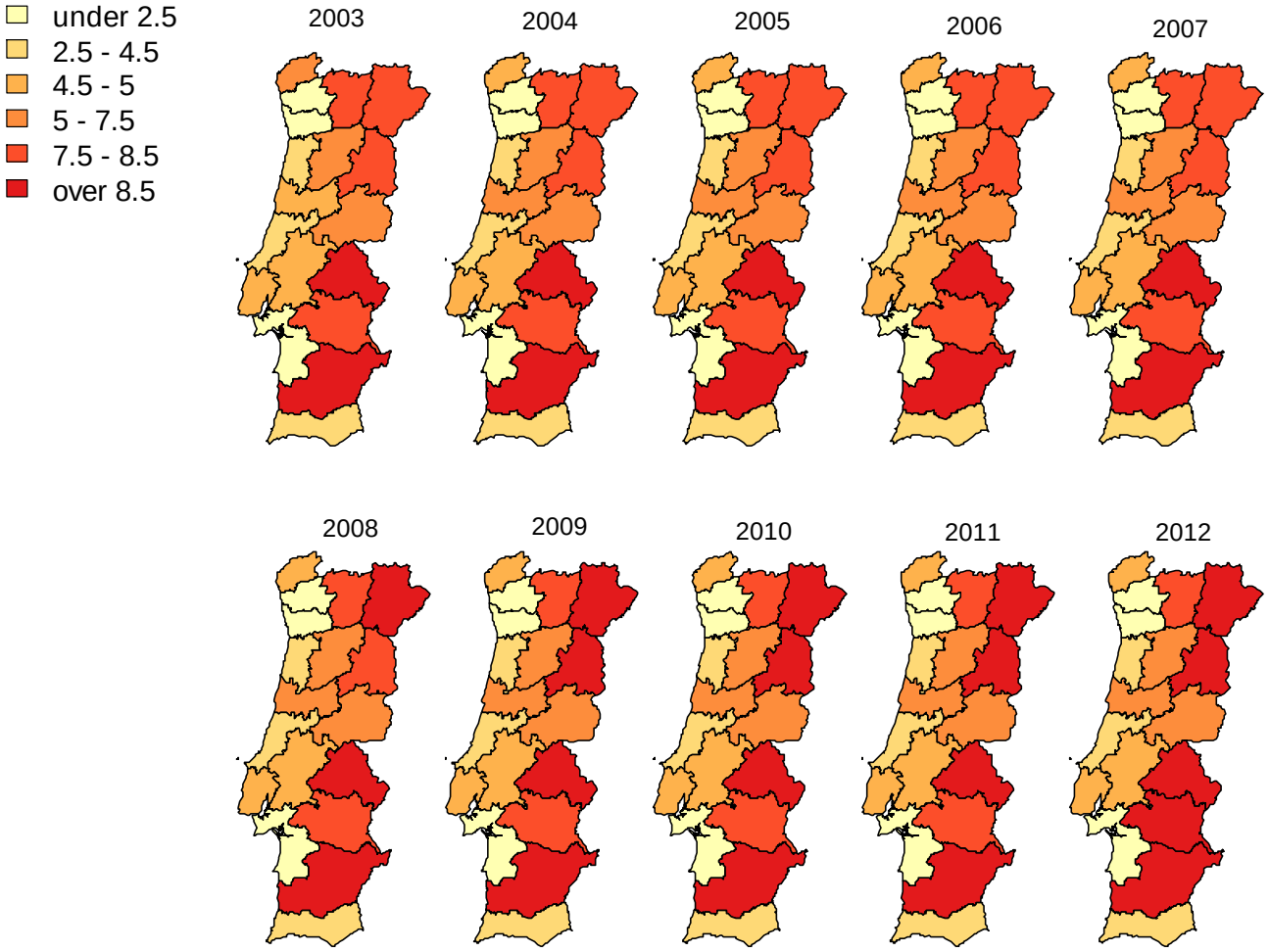


Figure 42. Boxplot with the number of primary care centres per 100,000 inhabitants in mainland Portugal, over time.



Oppositely to the hospitals, the highest number of primary care centres per 100,000 inhabitants is located in the east of mainland Portugal.

It was also in the east of Portugal that recorded the largest numbers of hospital admissions per 100,000 patients.

Figure 43. Maps representing the number of primary care centres per 100,000 inhabitants per district in mainland Portugal, over time.

4.5. MODELLING

Bayesian spatial-temporal models

A Bayesian spatial-temporal approach to estimate the annual number of hospital admissions due to heart failure was taken following the model structure described in section 3.3.1. using R2OpenBUGS.

We started by fitting a very simple model, with a fixed effect term for time and a structured error component.

```
model{
for(i in 1:N){
      for(k in 1:T){
            Y[i,k]~dpois(mu[i,k])
            log(mu[i,k])<-log(E[i,k])+alpha0+alpha1*tempo[k]+v[i]
      }
}
for(k in 1:T){
      tempo[k]<-k
}
for (i in 1:N){
      v[i]~dnorm(0,tau.v)
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
tau.v~dgamma(0.1,0.0001)
}
```

The estimates, obtained based on 25,000 iterations after a burn in period of 5,000 iterations, were as follows

	mean	sd	val2.5pc	median	val97.5pc	sample
alpha0	-0.09249	0.1004	-0.271	-0.09602	0.05358	25000
alpha1	0.04164	0.001059	0.03421	0.04162	0.04907	25000
deviance	4839	6.199	4829	4839	4854	25000
tau.v	5.457	1.848	2.507	5.242	9.6	25000

Deviance information	Dbar	Dhat	DIC	pD=Dbar-Dhat
Y	4839	4820	4858	18.97
total	4839	4820	4858	18.97

The next steps consisted in the inclusion of region-related independent variables (IV) in the model with fixed effect terms but negative pD values were obtained for all the models, indicating the existence of strong prior-data conflicts.

The alternative consisted in considering a simpler error structure, namely dropping the spatial error structure, which led us to the GLMM.

Generalized linear mixed-effects models

Generalized linear mixed-effects models were used as an alternative approach to estimate the annual number of HA. Spatial heterogeneity was accounted by considering the inclusion of random effects for the region-related IV.

The region-related IV were selected according to the results of exploratory analysis, and entered the model respecting the following order:

1) Year
2) Proportion of population aged ≥65 (prop65)
3) Average monthly income (inc)
4) Primary care access (pcaccss)
5) Hospital access (haccess)

Year was the first co-variate to be considered because of the temporal nature of data and the interest of analysing the data from a temporal point of view. The variable was shifted from 2003–2012 to 1–10.

Proportion of population aged over 65 was the first socio-demographic variable to be included because its spatial-temporal pattern was the most similar to the pattern of the number of HA.

The proportion of women in the population aged over 65 was not included in the model for not presenting spatial differences and remaining approximately constant over the years.

After the population, the AMI was the co-variate that presented the most similar pattern to the hospital admissions spatial-temporal distribution.

Since the temporal pattern of access to primary care centres and hospitals remained approximately constant over time, these were the last covariates to be considered for inclusion in the model.

According to the priority above mentioned, each of the IV started entering the model as a fixed effect and then as a random effect. In both cases, log-likelihood ratio test was used to assess whether the model provided a significantly better fit than the previously fitted models. A significance level of 5% was considered in the selection process. Due to the reasons pointed in section 3.3.2., LRT results are to be interpreted cautiously when applied to compare nested models differing by a random effect term.

The sequence of studied models up to the final model is presented below:

Model 1 (m1) – year as fixed effect

Model 2 (m2) – year as fixed and random effect

Log-likelihood ratio to compare m1vs. m2: $p < 0.001$ → m2 was selected

Model 3 (m3) – prop65 as fixed effect and year as fixed and random effect

Log-likelihood ratio to compare m2 vs. m3: $p = 0.2741$ → m3 was rejected

Model 3a (m3a) – prop65 and year as fixed and random effects

Log-likelihood ratio to compare m2 vs. m3a: $p < 0.001$ → m3a was selected

Model 4 (m4) – inc as fixed effect; prop65 and year as fixed and random effects

Log-likelihood ratio to compare m3a vs. m4: $p < 0.001$ → m4 was selected

Model 4a (m4a) – inc, prop65 and year as fixed and random effects

Log-likelihood ratio to compare m4 vs. m4a: $p < 0.001$ → m4a was selected

Model 4b (m4b) – pcaccess as fixed effect; inc, prop65 and year as fixed and random effects

Log-likelihood ratio to compare m4a vs. m4b: $p = 0.06971$ → m4b was rejected

Model 4c (m4c) – pcaccess, inc, prop65 and year as fixed and random effects

The inclusion of paccess as a random effect caused convergence problems, hence the variable was discarded from the model selection process→ m4c was rejected

Model 4d (m4d) – haccess as fixed effect; inc, prop65 and year as fixed and random effects

Log-likelihood ratio to compare m4a vs. m4d: $p = 1$ → m4d was rejected

Model 4e (m4e) – haccess, inc, prop65 and year as random effects

Log-likelihood ratio to compare m4a vs. m4e: $p < 0.001$ → m4e was selected

Model 4e is our final model including year, proportion of population aged ≥65, average monthly income and hospital access as fixed and random effects:

```
> summary(m4e)

   Generalized linear mixed model fit by maximum likelihood (Laplace
   Approximation) [glmerMod]
   Family: poisson  ( log )
   Formula: ha ~ year + prop65 + inc + haccess + offset(log(esp)) +
       (1 + year + prop65 + inc + haccess | district)

        AIC      BIC   logLik deviance df.resid
       2785     2849    -1373     2745      160

   Scaled residuals:
      Min     1Q Median     3Q    Max
   -6.700 -1.112  0.052  1.044  8.252

   Random effects:
    Groups   Name        Variance Std.Dev. Corr
    distrito (Intercept) 647.7952 25.452
             year        0.1997  0.447    0.93
             prop65      1.1914  1.092   -0.97 -0.89
             inc         1.1809  1.087   -0.66 -0.81  0.51
             haccess     0.0643  0.253    0.41  0.56 -0.41 -0.57


   Number of obs: 180, groups:  district, 18

   Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
   (Intercept)  -1.8271     6.4138   -0.28     0.78
   year          0.0639     0.1113    0.57     0.57
   prop65        0.0750     0.2745    0.27     0.78
   inc          -0.0598     0.2695   -0.22     0.82
   haccess       0.0203     0.0668    0.30     0.76

   Correlation of Fixed Effects:
              (Intr)  year   prp65 inc
   year        0.925
   prop65     -0.973 -0.886
   inc        -0.660 -0.808  0.511
   haccess     0.408  0.543 -0.419 -0.531
```

In the model selection procedure, although the inclusion of most of the covariates was seen as non-significant, we decided to keep them in the model since our objective was to establish a functional form of dependence of the response variable

on the region-related IV and time. This principle was applied in particular regarding covariate prop65.

Based on the model coefficients we estimated the fixed effects associated to each IV, as follows:

$$\log(\lambda_{it}) = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, year_{it} + (\beta_2 + b_{2i})\, prop65_{it} + (\beta_3 + b_{3i})\, inc_{it} + (\beta_4 + b_{4i})\, haccess_{it}$$

$$\log(\lambda_{it}) = -1.8271 + b_{0i} + (0.0639 + b_{1i})\, year_{it} + (0.0750 + b_{2i})\, prop65_{it} + (-0.0598 + b_{3i})\, inc_{it} + (0.0203 + b_{4i})\, haccess_{it}$$

$$\log(\mu_{it}) = \log(e_{it}) + \log(\lambda_{it})$$

$$\mu_{it} = e_{it} \times \exp[-1.8271 + b_{0i} + (0.0639 + b_{1i})\, year_{it} + (0.0750 + b_{2i})\, prop65_{it} + (-0.0598 + b_{3i})\, inc_{it} + (0.0203 + b_{4i})\, haccess_{it}]$$

$$\mu_{it} = e_{it} \times e^{-1.8271} \times e^{0.0639\, year_{it}} \times e^{0.0750\, prop65_{it}} \times e^{-0.0598\, inc_{it}} \times e^{0.0203\, haccess_{it}}$$

The estimated fixed effects indicate that, in average, over the districts,

1) The number of HA due to HF increase by 7% per year → exp(0.0639)=1.07
   As expected the number of HA increase over time. The model results support the exploratory analysis that indicated an average annual increase of 5% in the number of HA.
2) An increase of 1% in the proportion of population aged ≥65 accounts for an increase of 8% in HA → exp(0.075)=1.08

As described in the literature, HF is the most common reason for hospital admission in people aged over 65. According to the exploratory analysis, the Portuguese population has been aging in the last years. From 2003 to 2012, the proportion of people aged over 65 increased 2.5%, accounting 19.5% of the population in 2012. Hence, the proportion of people aged over 65 proved to be a demographic factor with a significant impact on the number of HA.

3) The increase of 100€ in the monthly income represents an average decrease of 5.8% in HA → exp(-0.0598)=0.942; 1-0.942=0.058
   Economic factors influence the access to healthcare. The spatial-temporal distribution of AMI is opposite to the number of hospital admissions per 100,000 inhabitants, which supports the model results that the increase of AMI leads to an average decrease in the number of hospital admissions.
4) 1 more hospital per 100,000 inhabitants accounts for an increase of 2% in HA → exp(0.0203)=1.02
   An increase in the number of hospitals allows to cover a larger population increasing the access to healthcare. As expected, a higher number of hospitals results in an increase of HA.

These changes are conditional to all the other IV remaining unchanged and refer to the country as a whole.

The Estimated random effects accounted for spatial heterogeneity (the differences between districts) by introducing corrections around the fixed effects.

In order to access the goodness of fit of our model (m4e) it was compared with two other models, presented below,

- **Model 5** – a GLMM without random effects for the region-related IV:

$$\log(\lambda_{it}) = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, year_{it} + (\beta_2)\, prop65_{it} + (\beta_3)\, inc_{it} + (\beta_4)\, haccess_{it}$$

This model states that districts may differ in the way they evolve in time concerning the number of HA. More precisely, the model accommodates the possibility of different slopes for time for the districts. As for the proportion of population above 65 years old, average monthly income and hospital access, here we assume that the number of HA relates to these IVs in the same manner.

Despite model 5 considers the same region-related IV that model 4e, but without random effect, the meaning of its coefficients is different. The increase in the proportion of people aged over 65 and number of hospitals per 100,000 inhabitants account for a decrease in the number of HA, which is different from the results achieved with model 4e.

```
> summary(mefmist)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
 Family: poisson  ( log )
Formula: ha ~ year + prop65 + inc + haccess + offset(log(esp)) +
    (1 + year | district)
   Data: data

     AIC      BIC   logLik deviance df.resid
    3480     3506    -1732     3464      172

Scaled residuals:
   Min     1Q Median     3Q    Max
-8.765 -1.727 -0.008  1.802  8.392

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 district (Intercept) 0.36402  0.6033
          year         0.00319  0.0565   -0.65
Number of obs: 180, groups:  district, 18

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.5569     0.8078    3.17   0.0015 **
year          0.1087     0.0195    5.57  2.6e-08 ***
prop65       -0.0439     0.0326   -1.35   0.1779
inc          -0.2450     0.0337   -7.26  3.8e-13 ***
haccess      -0.0390     0.0129   -3.03   0.0024 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
           (Intr) year prop65 inc
year        0.610
prop65     -0.955 -0.628
inc        -0.739 -0.661  0.569
haccess     0.285  0.169 -0.346 -0.123
```

- **Model 6** – a fixed effects model (GEE):

$$\log(\lambda_{it}) = \beta_0 + \beta_1\, year_{it} + \beta_2\, prop65_{it} + \beta_3\, inc_{it} + \beta_4\, haccess_{it}$$

GEE models are population-average or marginal models, providing a regression approach for generalized linear models when the responses are not independent. They allow to make inferences about the population accounting for the within-subject correlation. In our case, GEE models account for the within-district dependence but cannot incorporate the heterogeneity between the districts. The above model is presented here just for comparison purposes regarding the models ability to fit to the data and their predictive performance.

Because of the temporal structure of the data, we chose the autoregressive correlation structure, more precisely AR1.

```
Call:
geeglm(formula = ha ~ year + prop65 + inc + haccess +
    offset(log(esp)), family = poisson(link = "log"), data = data,
    id = district, corstr = "ar1")
 Coefficients:
            Estimate Std.err  Wald Pr(>|W|)
(Intercept)  -0.2308  0.3148  0.54     0.46
year          0.0658  0.0106 38.50  5.5e-10 ***
prop65        0.0546  0.0118 21.37  3.8e-06 ***
inc          -0.1370  0.0243 31.65  1.8e-08 ***
haccess       0.0277  0.0228  1.48     0.22
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     43.6    6.37
Correlation: Structure = ar1  Link = identity
Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.827  0.0595
Number of clusters:   18   Maximum cluster size: 10
```

According to Figure 44, our model (m4e) is the one that best fits to the observed data (number of hospital admissions), followed by model 5, which does not consider random effects for the region-related IV. Model 6 that only considered fixed effects has the worst quality of adjustment to the observed data. These results indicate that the random effects, used in our fitted model (m4e), have a crucial role in its goodness of fit.

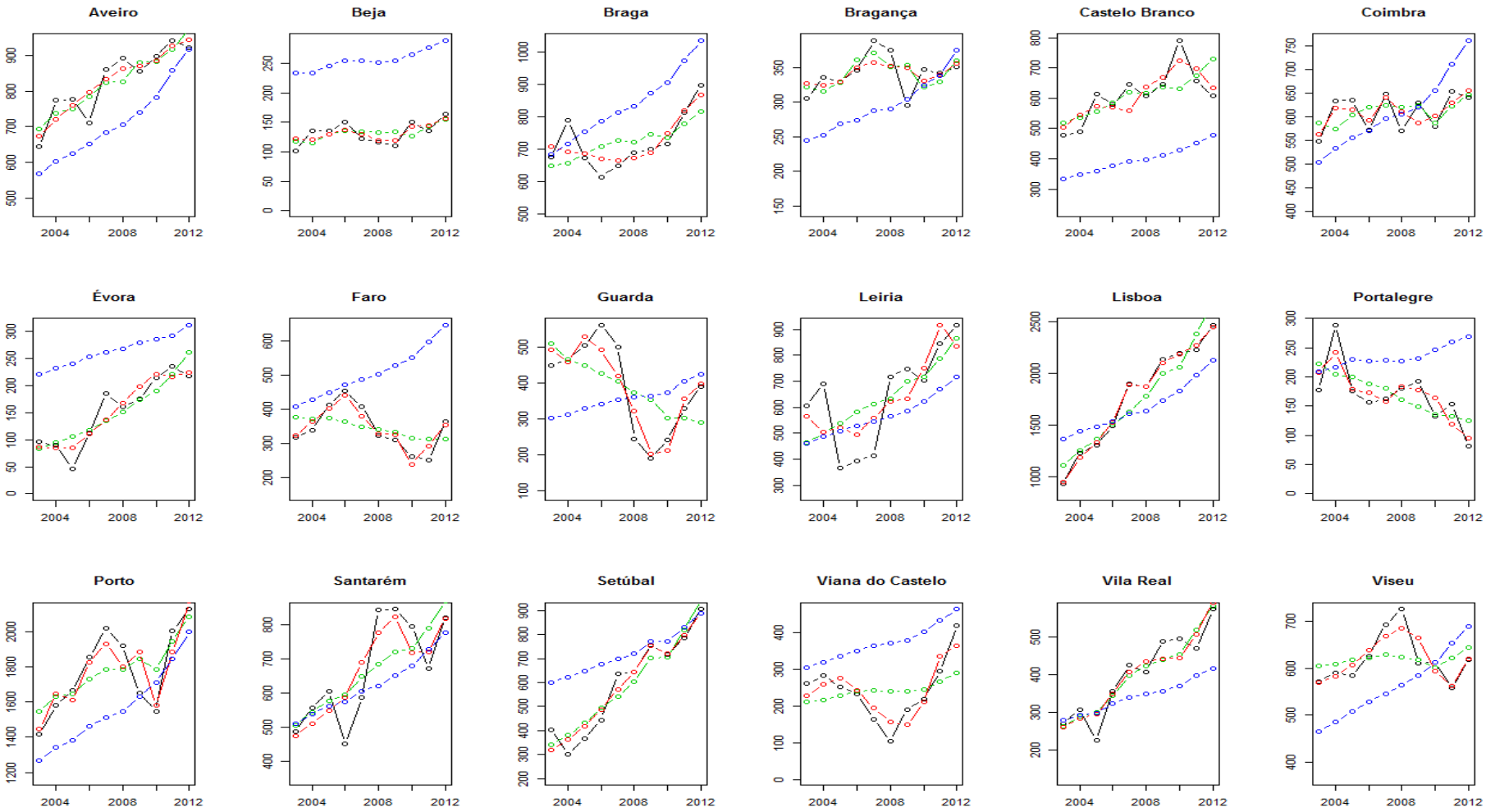


Figure 44. Observed and fitted number of hospital admissions: observed (black); fitted model – m4e (red); GLMM without random effects for the region-related IV – m5 (green); GEE – m6 (blue).

In addition to the graphical representation, mean absolute deviations (MAD) and the square root of mean squared errors (sqrtMSE) were calculated to assess goodness of fit of each model (Table 8). Model 4e has the lower MAD and MSE compared to the other models, supporting once again that this model fitted better.

Table 8. Mean absolute deviations (MAD) and the square root of mean squared error (sqrtMSE) of each model.

	Model 4e	Model 5	Model 6
MAD	34.7	56.5	131
sqrtMSE	49.3	77.5	160

Model performance and residual analysis

Fitted values were plotted against the observed values with the objective of assessing visually the goodness of fit. It is clearly seen that model m4e fits the data better than the other two models. The same can be seen when comparing model 5 with model 6, indicating that the inclusion of random effects provided a significant improvement in the ability of the model to describe the data (Figure 45; Figure 46; Figure 47).

For the residual analysis of each model, we computed the response (or raw) residuals, given by

$$r_{it} = y_{it} - \hat{\mu}_{it}$$

and the Pearson residuals, given by

$$r_{it}^{p} = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{\hat{\mu}_{it}}}.$$

Pearson residuals were plotted against the fitted values to check the existence of a non-random pattern. Residuals are expected to distribute homogenously (Figure 48; Figure 49; Figure 50).

In order to investigate the distribution of the residuals per district, the boxplots of the Pearson residuals were represented by district. In these representations, it can be seen that some districts show different dispersion values. These plots were

performed for the different models. The difference between the models is very clear. For the m4e model the residuals distribute very symmetrically around zero for all districts although showing different dispersion. For model 5, most of the districts still have the residuals symmetrically distributed around zero but districts of Guarda and Leiria show distinctive departures from zero. This aspect is further visible when considering model 6 (the GEE model), where all districts except Coimbra show residuals centred quite above or below zero (Figure 51). This is a very clear indication towards the use of random effects since departures from zero in the median of the residuals by district indicate that for some districts the values of HA are being underestimated whereas for the other these values are being overestimated.

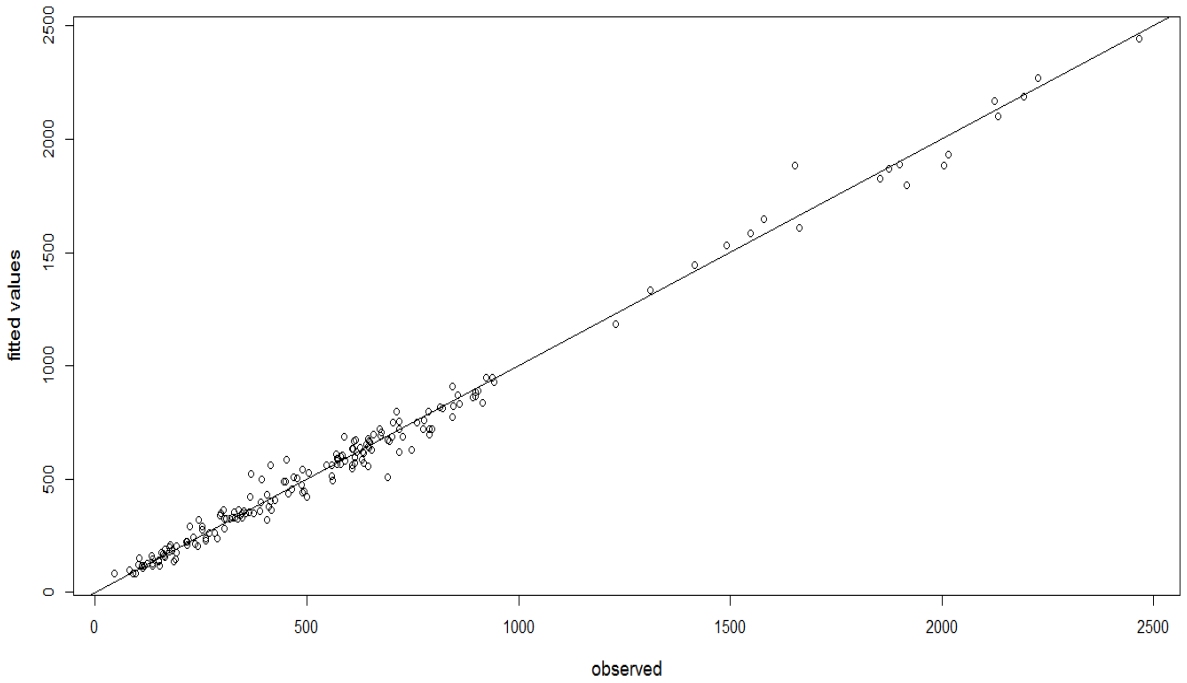


Figure 45. Model 4e: Fitted values vs. observed values.

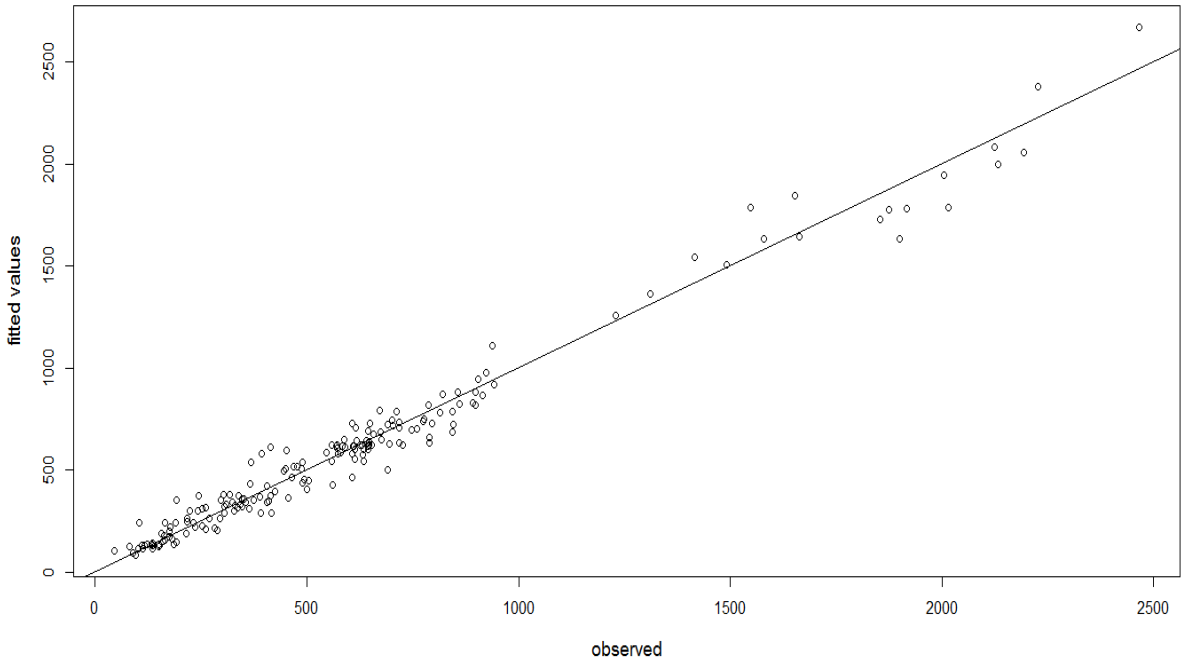


Figure 46. Model 5 (GLMM without random effects for region-related IV): Fitted values vs. observed values.

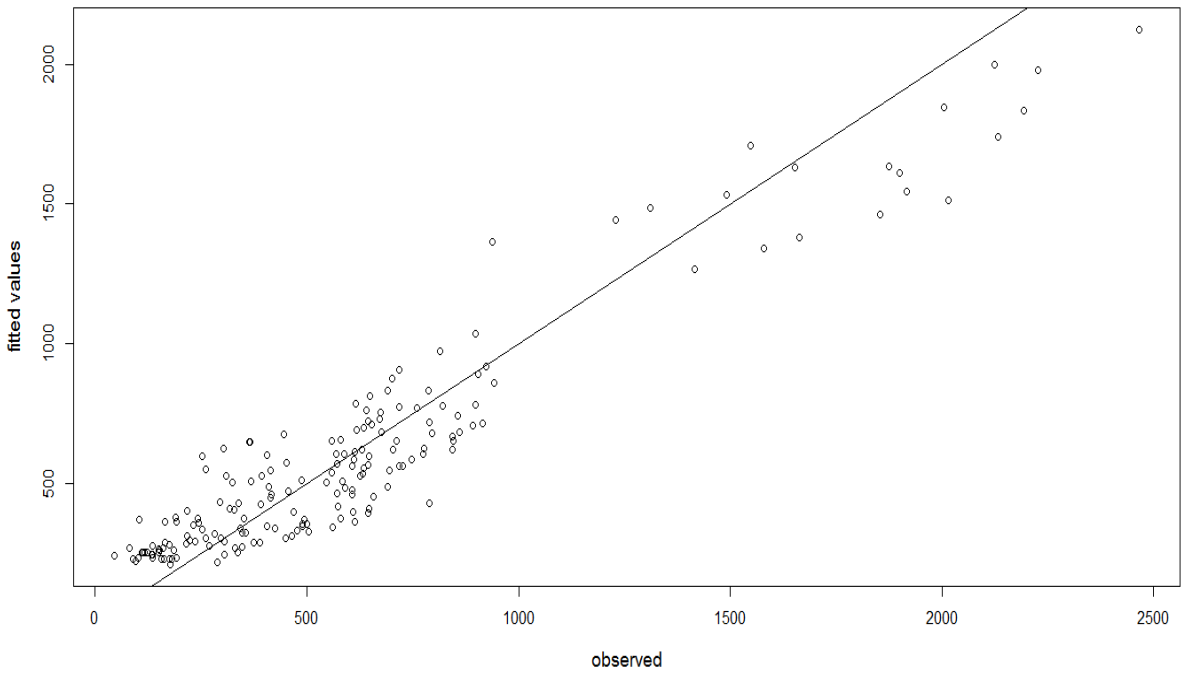


Figure 47. Model 6 (GEE): Fitted values vs. observed values.

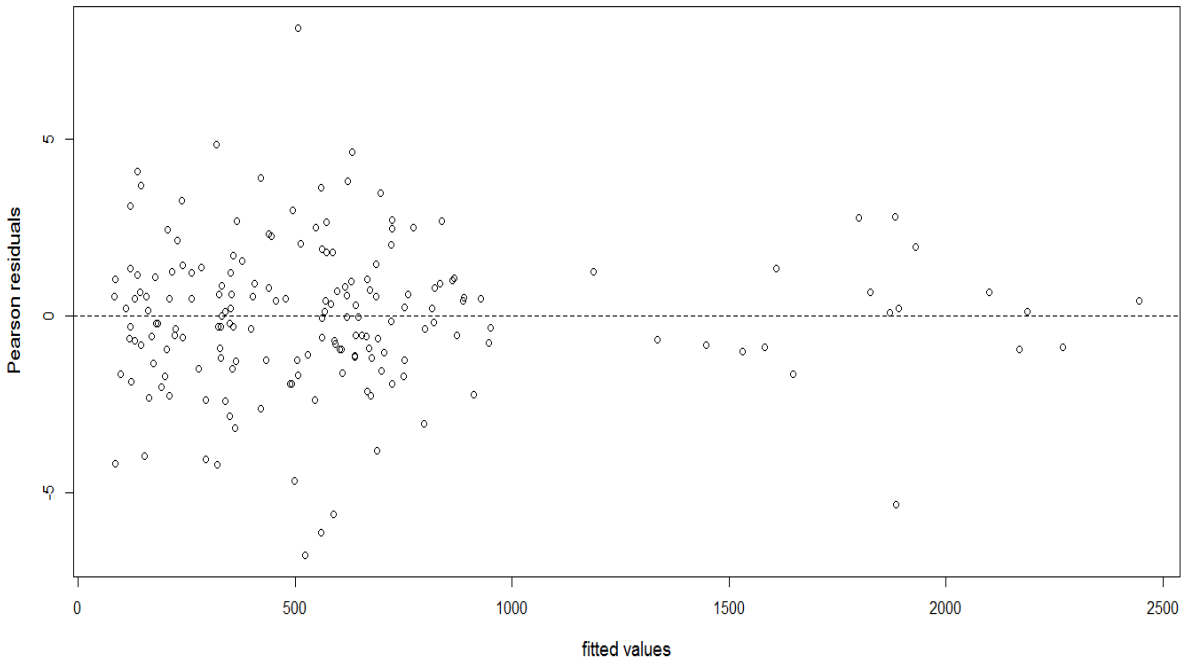


Figure 48. Model 4e: Pearson residuals vs. fitted values.

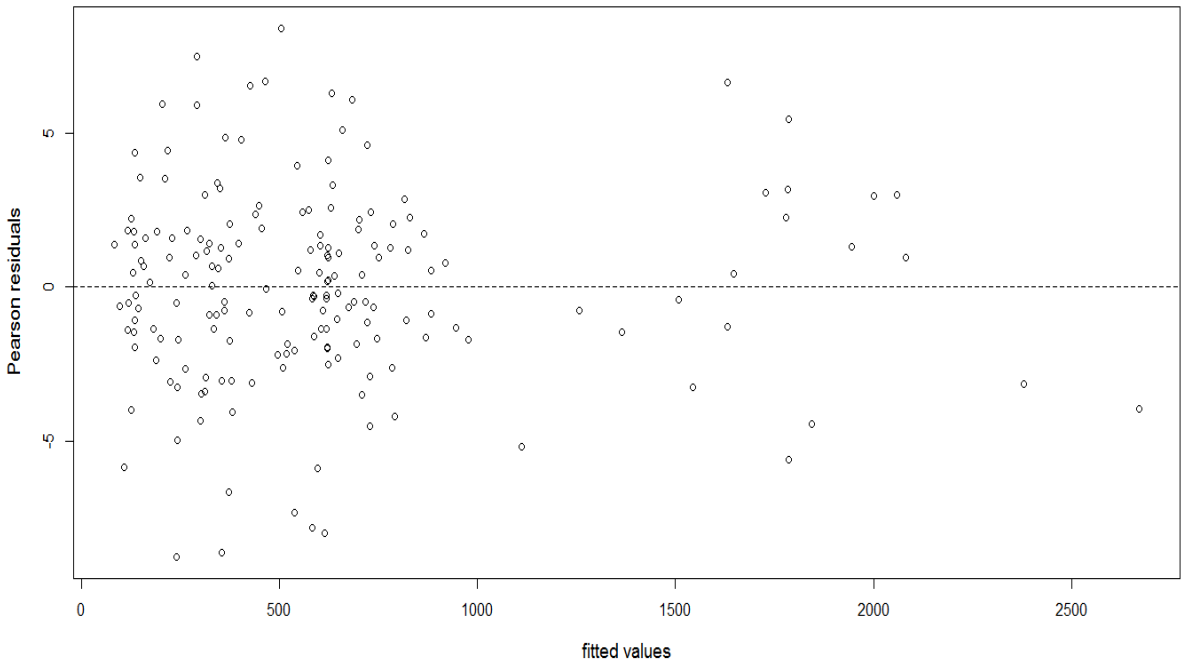


Figure 49. Model 5 (GLMM without random effects for region-related IV): Pearson residuals vs. fitted values.

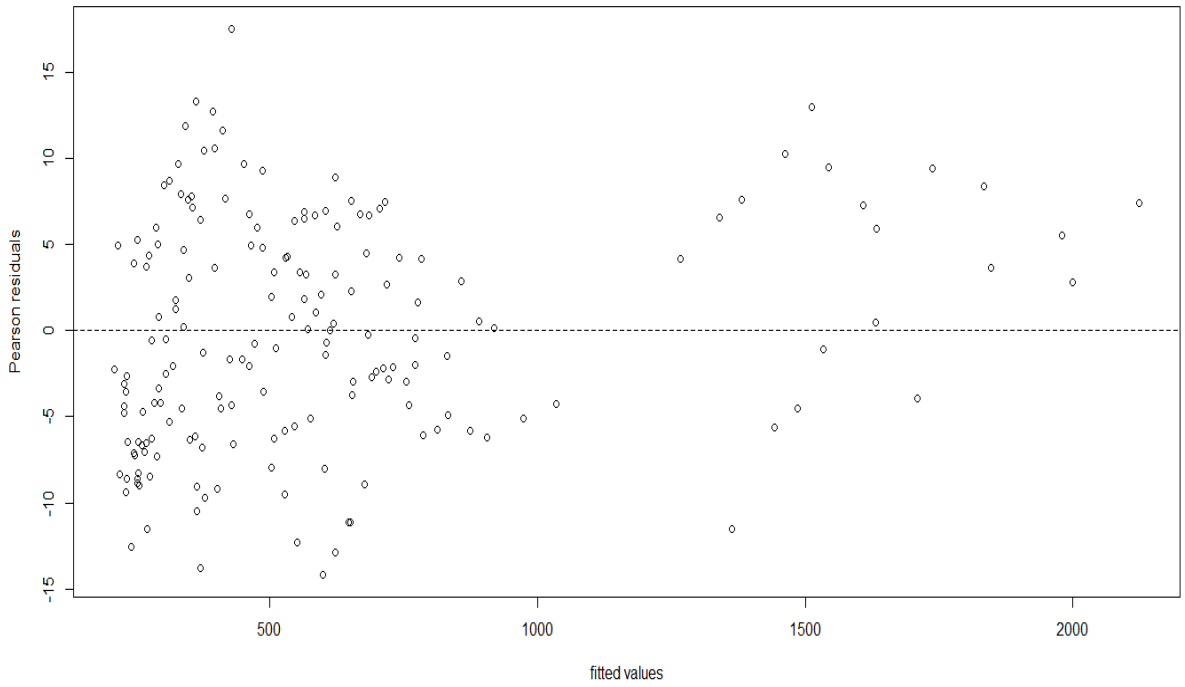


Figure 50. Model 6 (GEE): Pearson residuals vs. fitted values.

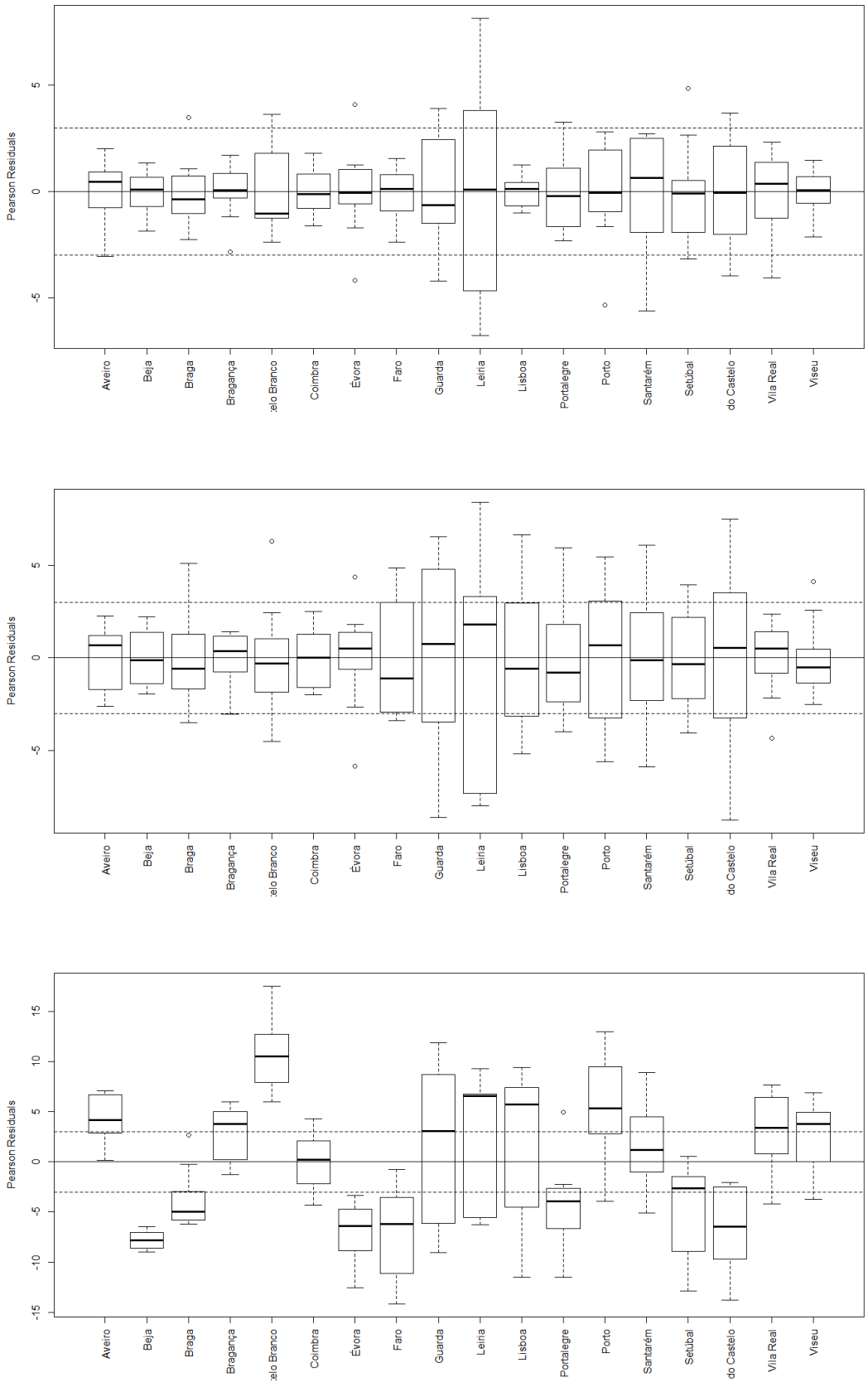


Figure 51. Residual analysis (from top to bottom): model 4e; model 5 (GLMM without random effects for the region-related IV); model 6 (GEE).

Predictive ability of the models

In order to assess the predictive ability of the models, a forecast for 2012 was performed based on 2003–2011 data. This procedure consisted in removing the data from 2012 and re-estimate the models. Then, `predict` function was used to obtain the predictions for the response variable giving as input the values for the IV for the year 2012. The predicted values were plotted against the observed values for 2012 for visual assessment of the predictive ability of the models. MAD and sqrtMSE were also calculated for a quantitative comparison. As expected, model 4e performed better than other models (Table 9).

Table 9. Mean absolute deviations (MAD) and the square root of mean squared error (sqrtMSE) of each model without 2012 data.

	Model 4e	Model 5	Model 6
MAD	80.4	104	110
sqrt(MSE)	113	133	156

Model 4e has the best predictive ability, with the predicted values quite close to the observed values. Leiria and Porto are the only districts with a prediction error above 200 (Figure 52), representing an overestimate of around 30% for Leiria (916 HA observed) and an underestimate around 20% for Porto (2125 HA observed).

Dropping the random effects from model 4e, resulting in model 5, did not affect significantly the predictive ability of the model. In this case, the districts with the higher prediction errors are Lisboa, Viana do Castelo and Castelo Branco (Figure 53).

Finally, the absence of random effects (model 6) affected the predictive ability of the model. Although the prediction errors are not higher in absolute value, a higher relative prediction errors were found for districts with lower number of HA, namely Faro, Beja and Portalegre, reaching values above 100% in some cases (Figure 54).

.

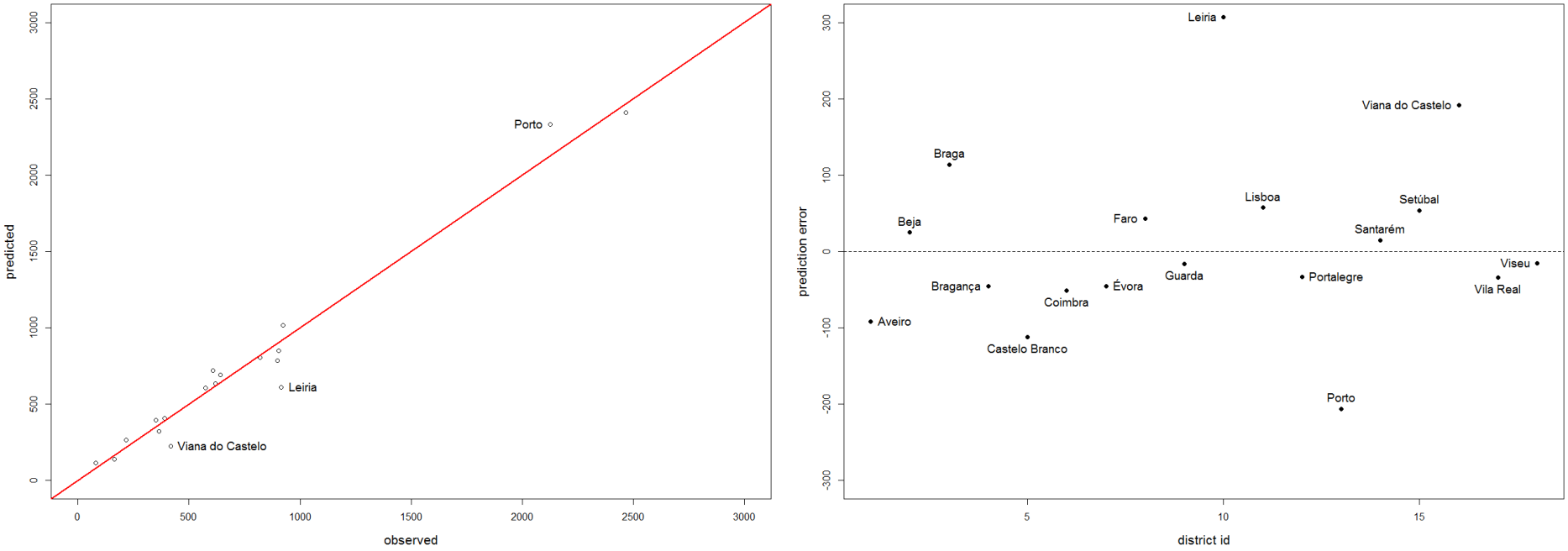


Figure 52. Model 4e – Left figure: observed values vs. fitted number of hospital admissions in 2012; Right figure: prediction error of the number of hospital admissions in 2012.

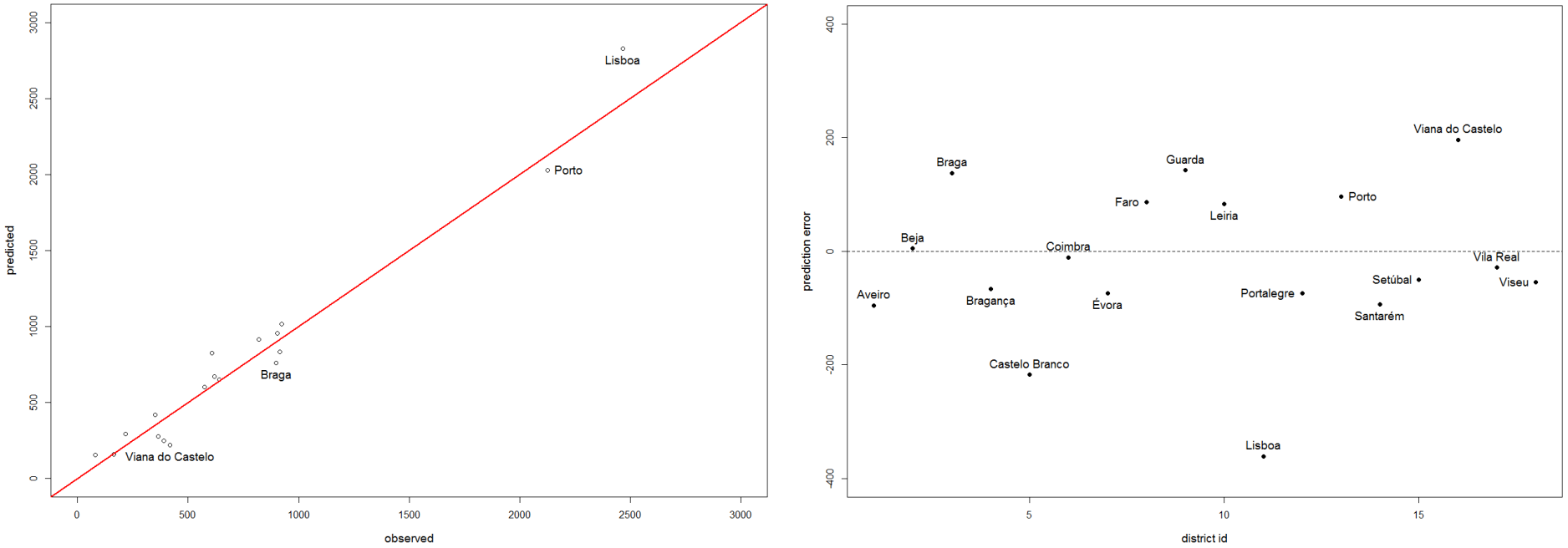


Figure 53. Model 5 (GLMM without random effects for region-related IV): – Left figure: observed values vs. fitted number of hospital admissions in 2012; Right figure: prediction error of the number of hospital admissions in 2012.

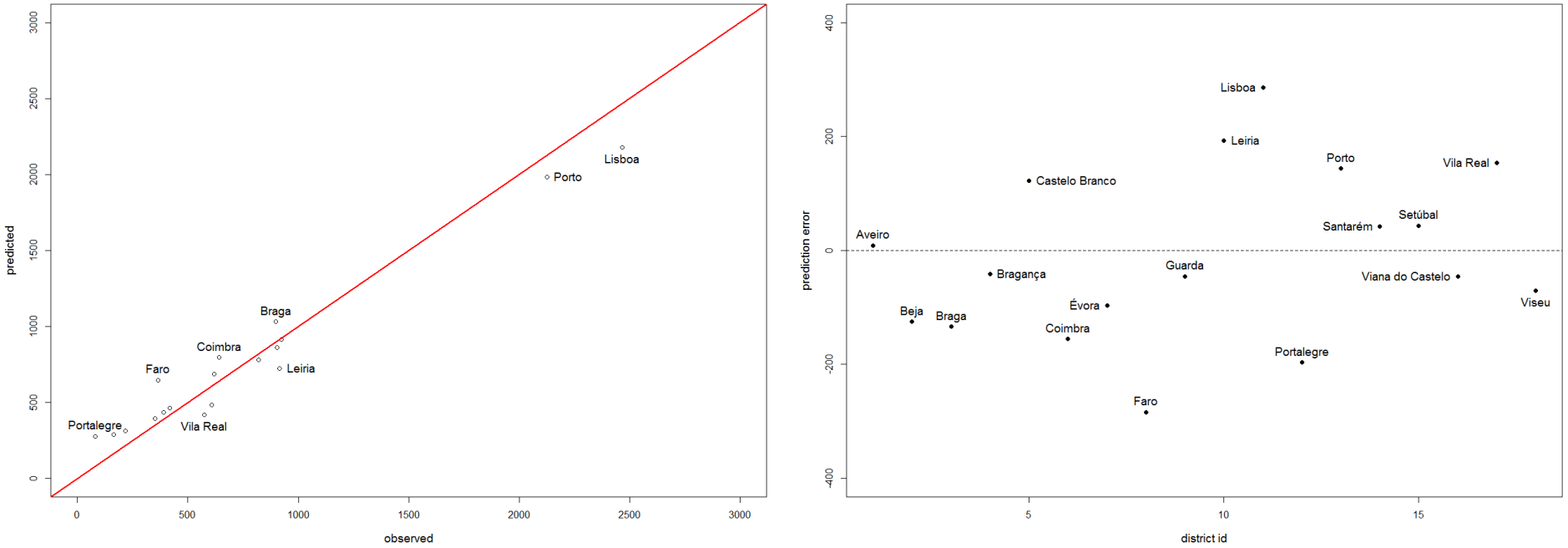


Figure 54. Model 6 (GEE) – Left figure: observed values vs. fitted number of hospital admissions in 2012; Right figure: prediction error of the number of hospital admissions in 2012.

5

DISCUSSION

Heart failure is complex syndrome with a multifactorial genesis characterized by symptoms such as breathlessness and fatigue, and signs such as fluid retention [1]. There are approximately 26 million people worldwide living with heart failure. The outlook for such patients is poor, with survival rates worse than those for bowel, breast or prostate cancer [**5**,**6**].

Heart failure is the most common reason for hospital admission in people aged over 65 years in economically developed regions, resulting in a significant burden for healthcare systems. Demand on healthcare services are predicted to increase over the next years as patient numbers rise due to the ageing populations, detrimental lifestyle changes and improved survival of those who can potentially develop HF as the final stage of another disease (e.g. hypertension, diabetes, stroke, etc.) [**16**]. Therefore, it becomes important to understand in which extent the environmental, demographic and socio-economic factors influence the number of hospital admissions, in order to identify areas of intervention that allow the development and implementation of public policies for heart failure.

The present work aimed to study the spatial-temporal pattern of HA in mainland Portugal and how external factors, like population characteristics, economics and access to healthcare infrastructures are influencing its evolution.

For that we used a dataset extracted from the national diagnosis-related group database, which contains the record of hospital admissions (with HF as the main diagnosis) per district of mainland Portugal, between 2003 and 2012. As covariates we considered: demographic factors (age and gender), economic factors (average monthly income) and social factors (number of hospitals and primary care centres).

The exploratory analysis, including the use of disease mapping techniques, revealed to be a very useful tool to characterize the spatial-temporal pattern of HA in Portugal. These tools also played an important role in the selection of relevant covariates to include in the modelling analysis. On the other hand, the modelling approach, especially the use of GLMM allowed to understand the effects of covariates in the number of HA and how these covariates could help to predict the number of HA.

As key results, we highlight:

- The number of HA due to HF have been increasing (35% between 2003 and 2012), with an estimated increase of 7% per year;
- The highest number of hospital admissions are located in the west of mainland Portugal;
- The Portuguese population has been aging in the last years, especially in the age group above 65 years, accounting for an increase of HA;
- Economic factors, like average monthly income seems to play a relevant role in the decrease of HA;
- The number of HA appears to increase as the hospital access increases (higher number of hospitals per inhabitant);
- Model 4e, including year, proportion of population aged ≥65, average monthly income and hospital access as fixed and random effects, was the model that fitted better to the observed data;
- Model 4e presented the best prediction ability compared to model 5 (without random effects for region-related IV) and model 6 (fixed effects).

The present work is, to our knowledge, the first study about the evolution of HA due to HF in Portugal. However, other studies have been conducted in similar areas. For example, the Portuguese study from Ceia F *et al*. that estimated the prevalence of chronic heart failure in mainland Portugal through a community-based epidemiological survey involving subjects attending primary care centres [**27**]. The authors found that the prevalence of CHF increases with age in both sexes, which is consistent with our results that indicated that a greater proportion of people aged above 65, resulted in an increase of HA.

Sayago-Silva I *et al.* 2013 studied the epidemiology of HF in Spain over the last 20 years, through a literature review. The authors concluded that the number of hospital admissions due to HF has increased over the last two decades, especially in people aged over 65 years, accounting for 3% of all hospital admissions and 2.5% of health care costs. These achievements are in line with our results, highlighting once again aging of population, as a predictive factor of HA increase [**46**].

Our study indicated that an increase in the number of hospitals per 100,000 inhabitants accounts for an increase of HA. This is consistent with Roemer's Law, a widely cited principle in health care policy, which states that hospital beds that are built tend to be used. Based on this principle, Delamater PL *et al.* 2013 investigated the effect of hospital bed availability on the utilization of hospital services, accounting for spatial structure and controlling for other known determinants of hospital utilization. The authors found compelling evidence that a positive, statistically significant relationship exists between hospital bed availability and inpatient hospitalization rates, independent of the changes in the geographic scale of analysis [**47**].

As previously mentioned, disease mapping is a powerful tool in the exploratory data analysis, providing a clear pattern of the study event (HA) and its covariates. The use of spatial-temporal maps to monitor diseases or health events is increasing due to the accuracy and ease of interpretation, enabling a prompt decision making by health experts. Research has been carried out to determine the potential of spatial-temporal maps in exploring disease and population patterns, revealing that these are effective techniques in identifying, interpreting and providing explanations about observed event with time and space distribution [**48**].

As an example of disease mapping application is the small-area studies, enabling the identification of disease distribution patterns, which become extremely popular in the field of public health. López-Abente G *et al.* 2014 used this technique to study the municipal spatial mortality patterns of the most frequent cancers in Spain, over a period of 20 years. As a result, the authors found that breast, colorectal and bladder cancer in women show signs of the possible spatial pattern that should be monitored [**49**].

In the present work, the results of the spatial and temporal analysis, supported by disease mapping exploratory techniques, guided the model building strategy. The differences found between the districts and the time evolution pattern supported the decision to use spatial-temporal models. Difficulties with the Bayesian approach constrained the scope of the analysis, limiting us to use models without a spatial structure (GLMM).

In the GLMM modelling analysis, the performance differences between the fitted models allowed to understand the impact of including random effect terms to account for the districts heterogeneity. Model 4e fitted better than model 5 (without random effects for region-related IV) showing that the number of HA due to HF depend on the covariates in a different manner from district to district. Concerning the predictive ability, no relevant differences were found between these two models indicating that the inclusion of random effects may be more relevant in explaining than in predicting. The statistical tools (namely the R packages for GLMM) should be further investigated for a more efficient use of its resources concerning estimation methods, model validation and inference.

Although proved useful, the GLMM class is not the most appropriate choice since none of these models incorporate any spatial dependence structure. Such feature could only be included in a Bayesian framework. Although the results can be considered satisfactory, we believe that Bayesian spatial-temporal models would render better results once estimation related problems were overcome. This idea is supported by the heterogeneity found in the graphical representation of the residuals by district, with some districts presenting Pearson residuals with much higher empirical variance than others. Bayesian spatial-temporal modelling should be further explored in future works.

6

CONCLUSIONS

This work, demonstrated that the evolution in the number of hospital admissions in Portugal can be substantially explained by time, aging population, the average monthly income and the access to hospital care. It was also demonstrated that the number of hospital admissions evolve differently in time depending on the districts, and that different model parameters are needed to explain such values in terms of the regional characteristics. Future work includes further investigation on the spatial dependency.

The conclusions of this work are consistent with results found in the literature, providing valuable information about the influence of external factors in the evolution of hospital admissions in Portugal.

7

BIBLIOGRAPHIC REFERENCES

1. Mcmurray, J.: ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Eur J Heart Fail 14, 803-69. (2012)

2. Yancy, C.: ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation 128, 240-327 (2013)

3. Dickstein, K. .: ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Eur Heart J 29, 2388-442 (2008)

4. Hunt, S.: Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 53, 1-90 (2009)

5. Bui, A. L. .: Epidemiology and risk profile of heart failure. Nat Rev Cardiol 8, 30-41 (2011)

6. Ambrosy, A. P. .: The global health and economic burden of hospitalizations for heart failure: lessons learned from HHF registries. J Am Coll Cardiol 63, 1123–33 (2014)

7. International Agency for Research on Cancer: Estimated cancer incidence, mortality and prevalence worldwide in 2012 (GLOBOCAN 2012) Lyon, 2012. (2012)

8. World Health Organization: Global Health Observatory: HIV/AIDS Geneva. (2014)

9. Lloyd-Jones, D. M. .: Lifetime risk for developing congestive heart failure: the Framingham Heart Study. Circulation 106, 3068–72 (2002)

10. Hodges, P.: Heart failure: epidemiologic update. Crit Care Nurs Q 32, 24-32 (2009)

11. Mozaffarian, D. .: Heart disease and stroke statistics-2015 update: a report from the american heart association. Circulation 131, 29-322 (2015)

12. Blair, J. E. .: Heart failure in North America. Curr Cardiol Rev 9, 128–46 (2013)

13. Al-Shamiri, M. Q.: Heart failure in the Middle East. Curr Cardiol Rev 9, 174–8 (2013)

14. Guo, Y. .: Heart failure in East Asia. Curr Cardiol Rev 9, 112–22 (2013)

15. Cowie, M. R. .: Improving care for patients with acute heart failure: before, during and after hospitalization., Oxford, UK (2014)

16. Ponikowski, P. .: Heart failure: preventing disease and death worldwide. (2014)

17. Khand, A. U. .: Do discharge codes underestimate hospitalisation due to heart failure? Validation study of hospital discharge coding for heart failure., 792–7 (2005)

18. Maggioni, A. P. .: EURObservational Research Programme: regional differences and 1-year follow-up results of the Heart Failure Pilot Survey (ESC-HF Pilot). Eur J Heart Fail 15, 808–17 (2013)

19. Tavazzi, L. .: Multicenter prospective observational study on acute and chronic heart failure: one-year follow-up results of IN-HF (Italian Network on Heart Failure) outcome registry. Circ Heart Fail 6, 473–81 (2013)

20. Nieminen, M. S. .: Gender related differences in patients presenting with acute heart failure. Results from EuroHeart Failure Survey II. Eur J Heart Fail 10, 140-8 (2008)

21. Sakata, Y. .: Gender differences in clinical characteristics, treatment and long-term outcome in patients with stage C/D heart failure in Japan. Circ J 2014 78, 428–35 (2014) Report from the CHART-2 study.

22. Bocchi, E. A. .: III Brazilian guidelines on chronic heart failure. Arq Bras Cardiol 93, 3–70 (2009)

23. Lloyd-Jones, D. .: Heart disease and stroke statistics – 2010 update: a report from the American Heart Association. Circulation 2010 121, 46-215 (2010)

24. Neumann, T. .: Heart failure: the commonest reason for hospital admission in

Germany: medical and economic perspectives. Dtsch Arztebl Int 106, 269–75 (2009)

25. World Health Organisation: Spending on health: A global overview (Fact sheet N°319). (2012)

26. World Health Organisation: General statistical procedures used to construct WHO health expenditure database. (2012)

27. Ceia, F. .: Prevalence of chronic heart failure in Southwestern Europe: the EPICA study. Eur J Heart Fail 4, 531–539 (2002)

28. Lawson, A. B. .: An Introductory Guide to Disease Mapping. Wiley-Blackwell, Chichester (2001)

29. Lawson, A. B. .: Disease Mapping with WinBUGS and MLwiN. Wiley-Blackwell, Chichester (2003)

30. Lawson, A. B.: Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology 2nd edn. CRC Press, Ltd, USA (2013)

31. Administração Central de Sistemas de Saúde (ACSS): Dados Nacional dos Grupos de Diagnóstico Homogéneos (GDH) de 2003 a 2012. (2014)

32. Instituto Nacional de Estatística (INE): População média anual residente (N.º) por Local de residência (Distrito/ Região), Sexo e Grupo etário (Por ciclos de vida); Anual - Estimativas Anuais da População Residente. (Accessed 2014) Available at: http://www.ine.pt

33. Instituto Nacional de Estatística (INE): Ganho médio mensal (€) por Localização geográfica (NUTS - 2002); Anual (1). (Accessed 2014) Available at: http://www.ine.pt

34. Instituto Nacional de Estatística (INE): Hospitais (N.º) por Localização geográfica e Natureza institucional; Anual - Inquérito aos Hospitais. (Accessed 2014) Available at: http://www.ine.pt

35. Instituto Nacional de Estatística (INE),: Centros de saúde (N.º) por Localização geográfica e Tipo de serviço; Anual - Inquérito aos Centros de Saúde. (Accessed 2014) Available at: http://www.ine.pt

36. Lawson, A. B.: Hierarchical modeling in spatial epidemiology. WIREs Computational Statistics 6, 405-417 (2014)

37. Mollié, A.: Bayesian mapping of disease. FL: Chapman and Hall/CRC Press, London, UK (1996)

38. Nobre, A. A. .: Spatio-temporal models for mapping the incidence of malaria in Pará. Environmetrics 16(3), 291-304 (2005)

39. Besag, J. .: Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43, 1-121 (1991)

40. Waller, L. A. .: Hierarchical Spatio-Temporal Mapping of Disease Rates. Journal of the American Statistical Association 92(438), 607-617 (1997)

41. Hedeker, D.: Generalized Linear Mixed Models. Encyclopedia of Statistics in Behavioral Science (2005)

42. Pinheiro, J. .: Mixed Effect Models in S and S-Plus., New York (2000)

43. Sun, L.: Comparison of Different Estimation Methods for Linear Mixed Models and Generalized Linear Mixed Models., Sweden (2011)

44. Bolker, B. .: GLMM for ecologists and evolutionary biologists. (Accessed July 26, 2013) Available at: http://glmm.wikidot.com/

45. Winter, B.: Linear models and linear mixed effects models in R: Tutorial 1. In: bodowinter.com. Available at: http://www.bodowinter.com/tutorial/bw_LME_tutorial1.pdf

46. Sayago-Silvaa, I. .: Epidemiology of Heart Failure in Spain Over the Last 20 Years. Rev Esp Cardiol 66(8), 649–656 (2013)

47. Delamater, P. L. .: Do More Hospital Beds Lead to Higher Hospitalization Rates? A Spatial Examination of Roemer's Law. PLoS One 8(2), e54900 (2013)

48. Ogao, P. J.: A tool for exploring space-time patterns : an animation user research. International Journal of Health Geographics 5(35), 1-8 (2006)

49. López-Abente, G.: Time trends in municipal distribution patterns of cancer mortality in Spain. BMC Cancer 14, 535 (2014)

8

APPENDIX

Appendix 1. Abstract of an oral vommunication presented *at I Encontro Luso-Galaico de Estatística em Ambiente e Ecologia (EES 2014)*.

I Encontro Luso-Galaico de Estatística em Ambiente e Ecologia (EES2014)

Comunicação Oral

Análise espaço-temporal dos internamentos hospitalares por insuficiência cardíaca em Portugal

Marília Antunes
Centro de Estatística e Aplicações e Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, marilia.antunes@fc.ul.pt

Daniel Viriato
Novartis Farma – Produtos Farmacêuticos S.A., Portugal, daniel.viriato@novartis.com

Palavras–chave: Disease Mapping, Modelos espaço-temporais, Insuficiência cardíaca.

Resumo: A Insuficiência Cardíaca (IC) é uma patologia com um elevado impacto económico e social a nível mundial. Apesar da existência de diversas alternativas terapêuticas para o tratamento da IC, não é clara qual a evolução real desta doença. Estudos internacionais apontam que o internamento por IC pode estar a diminuir, no entanto este pode ser fortemente condicionado por fatores ambientais, económicos e sociais. Com base em dados demográficos do Instituto Nacional de Estatística (INE) e na Base de Dados Nacional dos Grupos de Diagnóstico Homogéneos (GDH) da Administração Central de Sistemas de Saúde (ACSS) pretende-se avaliar de que forma o número de internamentos hospitalares, que tiveram como diagnóstico principal IC, tem variado nos vários distritos de Portugal Continental ao longo do tempo, bem como a associação existente entre os internamentos e as características demográficas e sócio-económicas que caracterizam as diferentes regiões. Para este efeito serão utilizadas técnicas de *Disease Mapping*, através da construção de modelos espaço-temporais, seguindo uma abordagem bayesiana.

Agradecimentos

Os autores agradecem à IASIST Portugal a disponibilização da base de dados com a informação agregada dos Grupos de Diagnóstico Homogéneos (GDH) da Administração Central de Sistemas de Saúde (ACSS).
Este trabalho foi parcialmente financiado pelos projectos:
PEst-OE/MAT/UI0006/2014 e PTDC/MAT/11835/2010.

Referências

[1] Ceia F et al. (2002) Prevalence of chronic heart failure in Southwestern Europe: the EPICA study. *Eur J Heart Fail* 4(4):531-9.

[2] Kim H et al. (2010) Comparison of Bayesian Spatio-Temporal Models for Chronic Diseases. *Journal of Data Science* 8:189-211.

[3] Lawson AB, Williams FLR. *An Introductory Guide to Disease Mapping*, John Wiley & Sons, Ltd, 2001. England.

[4] Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, 2nd Edition, CRC Press, Ltd, 2013. USA.

[5] Lawson AB, Browne WJ, Vidal Rodero, CL. *Disease Mapping with WinBUGS ans MLwiN*, John Wiley & Sons, Ltd, 2003. England.

Appendix 2. Abstract accepted for poster publication at ISPOR 18th Annual European Congress.

DISEASE MAPPING AND SPATIAL-TEMPORAL ANALYSIS OF HOSPITAL ADMISSIONS DUE TO HEART FAILURE IN PORTUGAL

Daniel Viriato[1], Marília Antunes[2]

[1]Novartis Farma – Produtos Farmacêuticos S.A., Porto Salvo, Portugal; [2]Faculdade de Ciências da Universidade de Lisboa-Departamento de Estatística e Investigação Operacional and CEAUL

OBJECTIVE: Heart failure (HF) is a major reason for hospital admissions (HA), with a high socio-economic impact. Therefore it is of utmost importance to understand how HA due to HF are evolving. This study aimed to build a predictive model for the annual number of HA due to HF in Portugal.

METHODS: Hospital admissions due to HF, between 2003 and 2012, were extracted from National Diagnosis-related group database. Demographic and socioeconomic data were collected per district, from Statistics Portugal. Generalized linear mixed-effects models (GLMM) were used to estimate the annual number of HA. Spatial heterogeneity was corrected by considering region-related independent variables (IV): proportion of population aged ≥65, average monthly income and hospital access. Random effects were considered for all IV.

RESULTS: The fixed effect estimates indicate that, in average, the number of HA due to HF increase by 7% per year. An increase of 1% in the proportion of population aged ≥65 accounts for an increase of 8% in HA. The increase of 100€ in the monthly income represents an average decrease of 5.8% in HA. By its turn, 1 more hospital per 100,000 inhabitants accounts for an increase of 2% in HA. These changes are conditional to all the other IV remaining unchanged. Estimated random effects accounted for spatial heterogeneity by introducing corrections around the fixed effects. The fitted model was compared to a GLMM without random effects for the region-related IV and a fixed effects model. Mean absolute deviations (MAD), used to assess goodness of fit, were 34.7, 56.5 and 131, respectively. Graphical representation also demonstrated that our model fitted better. Predictive ability of the model was assessed by MAD of forecast for 2012 based on 2003-2011 data (MAD=81).

CONCLUSIONS: Although this approach produced good results, the predictive ability could be further improved by the inclusion of other region-related variables.

Appendix 3. Total number of hospital admissions per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	645	775	777	711	860	892	856	899	942	923
Beja	102	135	135	151	122	116	111	151	135	164
Braga	677	789	674	615	649	691	701	718	815	898
Bragança	306	336	330	346	389	375	296	347	342	351
Castelo Branco	477	489	614	579	645	608	646	790	657	607
Coimbra	547	633	635	572	647	570	629	579	653	641
Évora	96	89	46	112	185	161	175	214	235	218
Faro	318	339	413	455	409	324	309	262	252	364
Guarda	449	465	504	561	500	244	191	242	328	391
Leiria	606	691	367	394	415	717	747	703	844	916
Lisboa	939	1230	1311	1491	1900	1874	2132	2194	2227	2466
Portalegre	177	289	176	156	162	181	192	133	153	81
Porto	1416	1580	1663	1854	2016	1917	1652	1548	2004	2125
Santarém	488	558	607	452	589	843	845	796	672	821
Setúbal	405	302	366	445	635	644	759	717	788	905
Viana do Castelo	261	282	252	232	164	104	190	218	295	417
Vila Real	269	306	224	355	424	406	489	494	469	573
Viseu	572	591	584	626	694	726	611	613	559	618
PORTUGAL	8750	9879	9678	10107	11405	11393	11531	11618	12370	13479

Appendix 4. Number of hospital admissions per 100,000 inhabitants, per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	89,85	107,87	108,13	98,92	119,65	124,19	119,36	125,61	131,94	129,72
Beja	63,86	84,94	85,38	96,01	77,94	74,55	71,79	98,29	88,53	108,46
Braga	80,49	93,51	79,67	72,52	76,37	81,24	82,43	84,52	95,99	105,91
Bragança	208,52	230,79	228,66	241,93	274,38	267,13	213,05	252,51	251,62	261,37
Castelo Branco	231,35	238,45	301,13	285,68	319,99	303,57	324,77	399,94	335,96	314,48
Coimbra	124,23	144,09	144,89	130,84	148,40	131,12	145,14	134,09	152,24	150,90
Évora	55,57	51,66	26,79	65,49	108,58	94,97	103,82	127,73	141,28	132,25
Faro	77,41	81,48	98,09	106,77	94,75	74,09	69,76	58,41	56,16	81,75
Guarda	254,58	266,50	291,96	328,60	296,22	146,33	115,99	148,90	204,65	247,59
Leiria	130,36	148,19	78,52	84,09	88,35	152,37	158,55	149,09	179,33	195,52
Lisboa	43,24	56,36	59,82	67,74	85,90	84,29	95,39	97,68	98,89	109,66
Portalegre	141,00	231,69	142,03	126,76	132,54	149,21	159,51	111,41	129,54	69,52
Porto	78,64	87,48	91,84	102,14	110,82	105,28	90,72	85,06	110,24	117,24
Santarém	106,97	122,29	133,02	99,08	129,09	184,84	185,48	174,99	148,24	182,13
Setúbal	50,19	37,14	44,70	53,97	76,44	76,94	90,02	84,45	92,34	105,91
Viana do Castelo	104,13	112,66	100,85	93,06	65,93	41,93	76,89	88,62	120,54	171,50
Vila Real	121,60	139,19	102,67	164,03	197,49	190,90	232,29	237,17	227,24	280,26
Viseu	145,54	151,00	149,89	161,41	179,76	189,02	159,98	161,50	148,30	165,24
PORTUGAL	87,83	98,94	96,77	100,90	113,66	113,39	114,67	115,50	123,15	134,74

Appendix 5. Resident population (in million inhabitants) per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	0,72	0,72	0,72	0,72	0,72	0,72	0,72	0,72	0,71	0,71
Beja	0,16	0,16	0,16	0,16	0,16	0,16	0,15	0,15	0,15	0,15
Braga	0,84	0,84	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85
Bragança	0,15	0,15	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,13
Castelo Branco	0,21	0,21	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,19
Coimbra	0,44	0,44	0,44	0,44	0,44	0,43	0,43	0,43	0,43	0,42
Évora	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,16
Faro	0,41	0,42	0,42	0,43	0,43	0,44	0,44	0,45	0,45	0,45
Guarda	0,18	0,17	0,17	0,17	0,17	0,17	0,16	0,16	0,16	0,16
Leiria	0,46	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47
Lisboa	2,17	2,18	2,19	2,20	2,21	2,22	2,24	2,25	2,25	2,25
Portalegre	0,13	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
Porto	1,80	1,81	1,81	1,82	1,82	1,82	1,82	1,82	1,82	1,81
Santarém	0,46	0,46	0,46	0,46	0,46	0,46	0,46	0,45	0,45	0,45
Setúbal	0,81	0,81	0,82	0,82	0,83	0,84	0,84	0,85	0,85	0,85
Viana do Castelo	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,24	0,24
Vila Real	0,22	0,22	0,22	0,22	0,21	0,21	0,21	0,21	0,21	0,20
Viseu	0,39	0,39	0,39	0,39	0,39	0,38	0,38	0,38	0,38	0,37
PORTUGAL	9,96	9,98	10,00	10,02	10,03	10,05	10,06	10,06	10,04	10,00

Appendix 6. Proportion of population aged over 65 years old per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	0,149	0,152	0,154	0,157	0,160	0,163	0,166	0,170	0,175	0,179
Beja	0,246	0,247	0,248	0,249	0,248	0,247	0,247	0,248	0,247	0,246
Braga	0,122	0,125	0,127	0,128	0,130	0,132	0,135	0,138	0,142	0,145
Bragança	0,251	0,255	0,259	0,263	0,266	0,270	0,275	0,280	0,283	0,284
Castelo Branco	0,258	0,260	0,262	0,263	0,264	0,265	0,266	0,268	0,270	0,271
Coimbra	0,201	0,204	0,206	0,209	0,211	0,212	0,215	0,218	0,223	0,227
Évora	0,231	0,233	0,236	0,236	0,237	0,237	0,238	0,239	0,241	0,243
Faro	0,187	0,187	0,187	0,187	0,187	0,187	0,188	0,189	0,193	0,197
Guarda	0,258	0,261	0,264	0,265	0,266	0,268	0,270	0,273	0,277	0,279
Leiria	0,181	0,183	0,185	0,187	0,189	0,191	0,193	0,196	0,201	0,204
Lisboa	0,165	0,167	0,170	0,173	0,175	0,178	0,182	0,186	0,191	0,195
Portalegre	0,265	0,266	0,267	0,266	0,266	0,265	0,265	0,265	0,265	0,265
Porto	0,129	0,131	0,134	0,136	0,138	0,141	0,144	0,149	0,154	0,158
Santarém	0,213	0,215	0,217	0,219	0,220	0,222	0,223	0,226	0,229	0,231
Setúbal	0,155	0,158	0,160	0,163	0,165	0,168	0,171	0,176	0,181	0,186
Viana do Castelo	0,205	0,208	0,211	0,213	0,214	0,216	0,219	0,223	0,226	0,228
Vila Real	0,209	0,213	0,216	0,219	0,222	0,225	0,229	0,233	0,237	0,239
Viseu	0,200	0,203	0,205	0,208	0,210	0,212	0,215	0,219	0,223	0,225
PORTUGAL	0,170	0,173	0,175	0,177	0,179	0,181	0,184	0,187	0,191	0,195

Appendix 7. Proportion of women in the population aged over 65 years old per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	0,578	0,579	0,578	0,578	0,578	0,578	0,578	0,579	0,579	0,579
Beja	0,559	0,561	0,564	0,567	0,569	0,572	0,575	0,578	0,580	0,582
Braga	0,587	0,588	0,588	0,589	0,589	0,588	0,588	0,588	0,588	0,588
Bragança	0,556	0,558	0,559	0,560	0,563	0,564	0,565	0,568	0,570	0,570
Castelo Branco	0,579	0,580	0,583	0,586	0,588	0,589	0,591	0,592	0,592	0,592
Coimbra	0,587	0,587	0,588	0,589	0,590	0,591	0,592	0,592	0,592	0,592
Évora	0,564	0,566	0,567	0,569	0,572	0,575	0,577	0,579	0,580	0,581
Faro	0,557	0,557	0,557	0,557	0,558	0,559	0,560	0,562	0,563	0,564
Guarda	0,580	0,581	0,582	0,583	0,586	0,587	0,588	0,590	0,592	0,593
Leiria	0,568	0,568	0,569	0,570	0,571	0,572	0,573	0,575	0,577	0,578
Lisboa	0,598	0,597	0,596	0,595	0,594	0,593	0,592	0,591	0,591	0,591
Portalegre	0,562	0,564	0,567	0,570	0,573	0,575	0,577	0,578	0,579	0,579
Porto	0,596	0,594	0,593	0,592	0,591	0,590	0,589	0,588	0,587	0,585
Santarém	0,581	0,581	0,582	0,583	0,585	0,585	0,586	0,588	0,589	0,589
Setúbal	0,568	0,567	0,567	0,567	0,567	0,567	0,566	0,567	0,568	0,568
Viana do Castelo	0,602	0,602	0,603	0,603	0,604	0,605	0,605	0,606	0,608	0,610
Vila Real	0,572	0,574	0,576	0,578	0,580	0,581	0,583	0,585	0,586	0,586
Viseu	0,578	0,578	0,579	0,580	0,581	0,582	0,583	0,585	0,586	0,587
PORTUGAL	0,583	0,583	0,583	0,583	0,584	0,584	0,584	0,584	0,585	0,585

Appendix 8. Average monthly income of the Portuguese population per district, over time.

	2003*	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	749	763	796	821	842	879	896	927	937	948
Beja	681	725	724	747	790	836	872	913	911	917
Braga	627	656	677	703	732	773	795	833	843	858
Bragança	616	653	666	681	700	752	777	820	841	825
Castelo Branco	646	669	692	710	726	766	790	822	831	835
Coimbra	732	761	779	805	832	870	902	941	942	951
Évora	729	748	774	799	820	848	865	901	919	922
Faro	736	758	778	803	836	866	891	926	932	933
Guarda	620	651	662	680	701	731	759	805	800	809
Leiria	723	743	770	795	829	866	888	918	928	938
Lisboa	1053	1073	1111	1144	1177	1219	1240	1284	1293	1308
Portalegre	693	721	722	754	779	821	847	863	866	878
Porto	757	776	814	834	864	907	931	972	977	984
Santarém	737	755	780	818	834	864	888	923	933	941
Setúbal	800	839	872	900	943	984	1002	1076	1093	1111
Viana do Castelo	633	660	681	701	734	778	816	840	845	848
Vila Real	634	657	703	700	712	745	788	825	831	845
Viseu	645	669	690	715	741	772	801	831	841	852
PORTUGAL	854	880	909	936	965	1010	1036	1076	1085	1096

* Data for 2003 were not available. These values were estimated from the straight line slope built from the values of the following years.

Appendix 9. Number of hospitals per 100,000 inhabitants per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	1,7	1,7	1,7	1,5	1,4	1,4	1,1	1,4	2,0	2,0
Beja	1,3	1,3	0,6	0,6	0,6	0,6	0,6	2,0	1,3	1,3
Braga	1,9	2,1	2,1	2,1	2,0	2,0	2,0	2,1	2,1	2,1
Bragança	2,0	2,1	2,1	0,7	0,7	0,7	0,7	2,2	2,2	3,0
Castelo Branco	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,5	1,5	1,6
Coimbra	3,9	4,6	4,1	3,7	3,9	3,7	3,5	4,4	4,4	4,5
Évora	1,7	1,7	1,7	2,3	2,3	2,4	2,4	2,4	1,8	1,8
Faro	1,9	1,9	1,7	1,9	1,9	1,6	1,6	1,8	2,0	2,0
Guarda	1,7	1,7	1,7	1,8	1,8	1,8	1,2	1,8	1,9	1,9
Leiria	2,4	2,4	2,4	2,1	2,1	2,3	1,9	2,5	2,5	2,6
Lisboa	5,3	5,3	5,2	4,9	5,2	4,7	4,7	5,3	5,2	5,2
Portalegre	3,2	3,2	3,2	2,4	1,6	1,6	1,7	2,5	2,5	2,6
Porto	1,9	1,9	1,9	2,0	1,9	1,8	1,9	2,4	2,1	2,2
Santarém	1,1	1,1	1,1	1,1	1,1	0,9	1,1	1,5	1,8	1,8
Setúbal	0,9	0,9	0,9	0,7	0,8	0,8	0,8	0,9	0,9	0,9
Viana do Castelo	0,4	0,4	0,4	0,4	0,8	0,8	0,8	1,2	1,2	1,2
Vila Real	0,9	0,9	0,9	1,4	0,9	0,9	1,0	1,4	1,5	1,5
Viseu	1,0	1,0	1,0	1,0	0,8	0,8	0,8	1,1	1,1	1,1
PORTUGAL	2,5	2,5	2,5	2,4	2,4	2,3	2,2	2,7	2,7	2,7

Appendix 10. Number of primary care centres per 100,000 inhabitants per district in mainland Portugal, over time.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Aveiro	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,7	2,7	2,7
Beja	8,8	8,8	8,9	8,9	8,9	9,0	9,1	9,1	9,2	9,3
Braga	2,3	1,8	1,8	1,8	1,8	1,8	1,8	1,8	2,2	2,2
Bragança	8,2	8,2	8,3	8,4	8,5	8,5	8,6	8,7	8,8	8,9
Castelo Branco	5,3	5,4	5,4	5,4	5,5	5,5	5,5	5,6	5,6	5,7
Coimbra	5,0	5,0	5,0	5,0	5,0	5,1	5,1	5,1	5,1	5,2
Évora	8,1	8,1	8,2	8,2	8,2	8,3	8,3	8,4	8,4	8,5
Faro	3,9	3,8	3,8	3,8	3,7	3,7	3,6	3,6	3,6	3,6
Guarda	7,9	8,0	8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,9
Leiria	3,7	3,6	3,6	3,6	3,6	3,6	3,6	3,6	3,6	3,6
Lisboa	4,6	4,5	4,5	4,5	4,5	4,5	4,4	4,4	4,4	4,4
Portalegre	12,7	12,8	12,9	13,0	13,1	13,2	13,3	13,4	13,5	13,7
Porto	2,6	1,9	2,0	2,0	1,9	1,9	1,9	1,9	2,4	2,4
Santarém	4,8	4,8	4,8	4,8	4,8	4,8	4,8	4,8	4,9	4,9
Setúbal	2,5	2,5	2,4	2,4	2,4	2,4	2,4	2,4	2,3	2,3
Viana do Castelo	5,2	4,8	4,8	4,8	4,8	4,8	4,9	4,9	4,9	4,9
Vila Real	7,2	7,3	7,3	7,4	7,5	7,5	7,6	7,7	7,8	7,8
Viseu	6,6	6,6	6,7	6,7	6,7	6,8	6,8	6,8	6,9	6,7
PORTUGAL	4,2	4,0	4,0	4,0	4,0	4,0	4,0	4,0	4,1	4,1

Appendix 11. Script used to build maps (R Statistics).

```
pop<-read.table("populacaodistritos.csv",header=T,dec=",",sep=";")

                Libraries for maps
library(sp)
library(maptools)
gpclibPermit()
library(RColorBrewer)
library(classInt)

                                    #
distritos_mapas <- readShapePoly("distritos_nao_militar.shp")
colours <- brewer.pal(7, "PuBu")

          Example: Resident population          #

populacao<-pop[,2]
for (i in 3:11)populacao<-c(populacao,pop[,i])
populacao
summary(populacao)
p<-c(1/6,2/6,3/6,4/6,5/6)
b<-quantile(populacao,p)
#brks<-c(0,b,2500000)
brks<-c(0,150000,175000,450000,550000,750000,1500000,2500000)
brksround<-round(brks,0)

distritos_mapas$pop2003<-pop[,2]
distritos_mapas$pop2004<-pop[,3]
distritos_mapas$pop2005<-pop[,4]
distritos_mapas$pop2006<-pop[,5]
distritos_mapas$pop2007<-pop[,6]
distritos_mapas$pop2008<-pop[,7]
distritos_mapas$pop2009<-pop[,8]
distritos_mapas$pop2010<-pop[,9]
distritos_mapas$pop2011<-pop[,10]
distritos_mapas$pop2012<-pop[,11]

par(mfrow=c(2,5),mar=c(0,0,0,0),oma = c(0, 0, 2, 0))
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2003,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2003",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2004,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2004",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2005,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2005",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2006,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2006",cex=1.5)
```

```
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2007,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2007",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2008,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2008",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2009,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2009",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2010,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2010",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2011,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2011",cex=1.5)
plot(distritos_mapas, col=colours[findInterval(distritos_mapas$pop2012,
brksround,all.inside=TRUE)], axes=F)
text(locator(1),"2012",cex=1.5)
#mtext("População residente 2003 - 2012", outer = TRUE, cex = 1.3)

#Legend
par(mfrow=c(1,1))
plot(c(0,0),col="white",axes=F,xlab="",ylab="")
legend(locator(1), legend=leglabs(brksround), fill=colours, bty="n",cex=1.5)
```

Appendix 12. Script used to estimate the values of average monthly income in 2003 (R Statistics).

```
data<-read.table("rendimentos.csv",head=T,sep=";",dec=",")
head(data)
nomes<-data$Distrito
dist<-rep(nomes,9)
dist<-sort(dist)
ano<-rep(c(2004:2012),19)
length(ano)
dados<-data.frame(dist)
dados$ano<-ano
rend<-matrix(0,nrow=19,ncol=9)
rend<-data[1:19,3:11]
rendimento<-vector()
rendimento<-as.vector(rend[1,])
for(i in 2:19)rendimento<-c(rendimento,rend[i,])
rendimento<-unlist(rendimento)
dados$rendimento<-rendimento
dados<-data.frame(dados)

write.table(dados,"RendDistrito.txt",col.names=T,row.names=F,quote=F,sep=";")

require(nlme)

data<-read.table("RendDistrito.txt",head=T,sep=";")
head(data)
summary(data)
m1.list<-lmList(rendimento~ano|dist,data=data)
sumario<-summary(m1.list)
sum4<-sumario[4]
mode(sum4)
coefs<-as.vector(unlist(sum4))
coefs

intercept<-coefs[1:19]
decl<-coefs[77:95]
prev<-intercept+decl*2003

data<-read.table("rendimentos.csv",head=T,sep=";",dec=",")
head(data)
data$ano2003<-round(prev,2)
data<-data[,c(1,12,3:11)]
names(data)<-c("distrito","2003","2004","2005","2006","2007","2008",
"2009","2010","2011","2012")
data

write.table(data,"RendCompleto.csv",col.names=T,row.names=F,quote=F,sep=";")
```

Appendix 13. Script used to build boxplots (R Statistics).

```
propint<-read.table("PropInternamentos.csv",head=T,sep=";",dec=",")
names(propint)<-c("distrito",c(2003:2012))
head(propint)

nomes<-propint$Distrito
dist<-rep(nomes,9)
dist<-sort(dist)

     # Example: Number of hospital admissions per 100,000 inhabitants       #

propintaux<-propint[,2]
for(i in 3:11)propintaux<-c(propintaux,propint[,i])
anos<-sort(rep(c(2003:2012),18))
aux<-data.frame(cbind(propintaux,anos))
popanual<-colSums(popdist[,2:11])
internamentosanual<-colSums(internamentos[,2:11])
taxapais<-internamentosanual/popanual*100000
media10anos<-sum(internamentosanual)/sum(popanual)*100000
boxplot(aux$propintaux~aux$anos, xlab="Time (years)",
ylab="Nr. of hospital admissions (per 100,000 inhabitants)", col="grey")
points(c(1:10),taxapais,col=4,type="b",lwd=2,pch=18)
abline(h=media10anos,col=2,lty=2,lwd=2)

text(1,270,"Guarda",cex=0.7)
text(2,280,"Guarda",cex=0.7)
text(3,315,"Castelo Branco",cex=0.7)
text(3,280,"Guarda",cex=0.7)
text(4,265,"Castelo Branco",cex=0.7)
text(4,340,"Guarda",cex=0.7)
text(7,340,"Castelo Branco",cex=0.7)
text(8,390,"Castelo Branco",cex=0.7)
text(9,350,"Castelo Branco",cex=0.7)

legend("topleft", inset=.01, c("Annual","10 year"), col=c(4,2),lwd=2,lty=c(1,2),
pch=c(18,30), horiz=FALSE, cex=0.75)
```

Appendix 14. Script used to build lattice xyplot (R Statistics).

```
# Hospital admissions
lattice::xyplot(int~ano | distrito, groups=distrito, data=data,
type=c('p','r'), auto.key=F,ylab="Nr. of hospital admissions",xlab="Time (years)")

# hospital admissions per 100,000 inhabitants
lattice::xyplot(int/pop~ano | distrito, groups=distrito, data=data,
type=c('p','r'), auto.key=F,ylab="Nr. of hospital admissions (per 100,000 inhabitants)",xlab="Time (years)")

# Resident population
lattice::xyplot(pop~ano | distrito, groups=distrito, data=data,
 type=c('p','r'), auto.key=F,xlab="Time (years)",ylab="Number of inhabitants (in million)")

# Population aged over 65
lattice::xyplot(popmais65~ano | distrito, groups=distrito, data=data,
 type=c('p','r'), auto.key=F,xlab="time",ylab="population 65+ (x 100000)")

# Proportion of population aged over 65
lattice::xyplot(propmais65~ano | distrito, groups=distrito, data=data1,
type=c('p','r'), auto.key=F,xlab="Time (years)",
ylab="Proportion of inhabitants aged over 65")

#  Proportion of women in the population aged over 65
lattice::xyplot(propmulh65~ano | distrito, groups=distrito, data=data1,
 type=c('p','r'), auto.key=F,xlab="Time (years)",
ylab="Proportion of women")

# Hospitals per 100,000 inhabitants
lattice::xyplot(acessohosp~ano | distrito, groups=distrito, data=data,
 type=c('p','r'), auto.key=F,xlab="Time (years)",
ylab="Hospitals per 100,000 inhabitants ")

# Primary care centres per 100,000 inhabitants
lattice::xyplot(acessocsaude~ano | distrito, groups=distrito, data=data,
 type=c('p','r'), auto.key=F,xlab="Time (years)",
ylab="Primary care centres per 100,000 inhabitants")

# Average monthly income
lattice::xyplot(rend~ano | distrito, groups=distrito, data=data,
 type=c('p','r'), auto.key=F,xlab="Time (years)",
ylab="Average monthly income (x100 €)")
```

Appendix 15. Bayesian modelling analysis (OpenBugs).

```
############  modelo 00
model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+u[i]
        }
}

u[1:N]~car.normal(adj[],weights[],num[],tau.u)
for(k in 1:SumNumNeigh){
        weights[k]<-1
        }

alpha0~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
}

############  modelo 01

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+u[i]+v[i]
        }
}
u[1:N]~car.normal(adj[],weights[],num[],tau.u)
for(k in 1:SumNumNeigh){
        weights[k]<-1
        }
for (i in 1:N){
        v[i]~dnorm(0,tau.v)
        }
alpha0~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.v~dgamma(0.1,0.0001)
}

############  modelo 02

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+alpha1*tempo[k]+u[i]+v[i]
        }
}
for (k in 1:T){
        tempo[k]<-k
}
u[1:N]~car.normal(adj[],weights[],num[],tau.u)
for(k in 1:SumNumNeigh){
        weights[k]<-1
        }
for (i in 1:N){
        v[i]~dnorm(0,tau.v)
```

```
        }
alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.v~dgamma(0.1,0.0001)
}


############  modelo 03

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+alpha1*tempo[k]+v[i]
        }
}
for(k in 1:T){
        tempo[k]<-k
}
for (i in 1:N){
v[i]~dnorm(0,tau.v)
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
tau.v~dgamma(0.1,0.0001)
}
                mean            sd              val2.5pc        median          val97.5pc
        sample
alpha0          -0.09249        0.1004          -0.271          -0.09602        0.05358
        25000
alpha1          0.04164                 0.001059        0.03421                 0.04162
        0.04907                 25000
deviance        4839.0          6.199           4829.0          4839.0          4854.0          25000
tau.v           5.457           1.848           2.507           5.242           9.6             25000
Deviance information
        Dbar    Dhat    DIC     pD
Y       4839.0  4820.0  4858.0  18.97
total   4839.0  4820.0  4858.0  18.97

############  modelo 04

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+alpha1*tempo[k]+alpha2*prop65[i,k]+ v[i]
        }
}
for(k in 1:T){
        tempo[k]<-k
}
for (i in 1:N){
        v[i]~dnorm(0,tau.v)
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
```

```
alpha2~dnorm(0,0.001)
tau.v~dgamma(0.1,0.0001)
}
```

	mean	sd	val2.5pc	median	val97.5pc	sample
alpha0	-2.842	0.2417	-3.128	-2.872	-2.372	25000
alpha1	0.003538	0.003297	-0.005184	0.003143	0.01325	25000
alpha2	13.95	1.147	11.74	13.91	15.74	25000
deviance	4650.0	6.819	4639.0	4650.0	4665.0	25000
tau.v	5.041	1.841	2.161	4.804	9.241	25000

Deviance information

	Dbar	Dhat	DIC	pD
Y	4650.0	4630.0	4670.0	20.05
total	4650.0	4630.0	4670.0	20.05

```
############  modelo 05

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+alpha1*tempo[k]+b1[i]*tempo[k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}


alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
}

############  modelo 06

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-
log(E[i,k])+alpha0+alpha1*tempo[k]+alpha2*prop65[i,k]+b1[i]*tempo[k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
```

```
}


alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
}

############  modelo 07
model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-
log(E[i,k])+alpha0+alpha1*tempo[k]+alpha2*prop65[i,k]+b1[i]*tempo[k]+b2[i]*prop65[i,k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
}


############  modelo 08

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+
                alpha1*tempo[k]+alpha2*prop65[i,k]+alpha3*mulhM65[i,k]+
                b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*mulhM65[i,k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
```

```
        tempo[k]<-k
}


alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
alpha3~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
}

###########  modelo 09

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+
                alpha1*tempo[k]+alpha2*prop65[i,k]+alpha3*hospit[i,k]+
                b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*hospit[i,k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
alpha3~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
}

###########  modelo 10

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+
                alpha1*tempo[k]+alpha2*prop65[i,k]+alpha3*hospit[i,k]+alpha4*rend[i,k]+
                b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*hospit[i,k]+b4[i]*rend[i,k]+u[i]
        }
}
for(i in 1:N){
```

```
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
        b4[i]~dnorm(0,tau.b4)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
alpha3~dnorm(0,0.001)
alpha4~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
tau.b4~dgamma(0.1,0.0001)
}

############  modelo 11

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+

        alpha1*tempo[k]+alpha2*prop65[i,k]+alpha3*rend[i,k]+alpha4*hospit[i,k]+alpha5*csaude[i,k]+
                b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*rend[i,k]+b4[i]*hospit[i,k]+b5[i]*csaude[i,k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
        b4[i]~dnorm(0,tau.b4)
        b5[i]~dnorm(0,tau.b5)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}


alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
alpha3~dnorm(0,0.001)
alpha4~dnorm(0,0.001)
alpha5~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
```

```
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
tau.b4~dgamma(0.1,0.0001)
tau.b5~dgamma(0.1,0.0001)
}
############  modelo 12

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+

        alpha1*tempo[k]+alpha2*prop65[i,k]+alpha3*rend[i,k]+alpha4*hospit[i,k]+alpha5*csaude[i,k]+al
pha6*mulhM65[i,k]+

        b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*rend[i,k]+b4[i]*hospit[i,k]+b5[i]*csaude[i,k]+b6[i]*mulhM65
[i,k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
        b4[i]~dnorm(0,tau.b4)
        b5[i]~dnorm(0,tau.b5)
        b6[i]~dnorm(0,tau.b6)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}


alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
alpha3~dnorm(0,0.001)
alpha4~dnorm(0,0.001)
alpha5~dnorm(0,0.001)
alpha6~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
tau.b4~dgamma(0.1,0.0001)
tau.b5~dgamma(0.1,0.0001)
tau.b6~dgamma(0.1,0.0001)
}

############  modelo 14
model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-
log(E[i,k])+b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*rend[i,k]+b4[i]*csaude[i,k]+u[i]
```

```
        }
}
for(k in 1:T){
        tempo[k]<-k
}
for (i in 1:N){
        u[i]~dnorm(0,tau.u)
        }
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
        b4[i]~dnorm(0,tau.b4)
}
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
tau.b4~dgamma(0.1,0.0001)
}

############  modelo 15
model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-
log(E[i,k])+b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*rend[i,k]+b4[i]*csaude[i,k]+u[i]
        }
}
for(k in 1:T){
        tempo[k]<-k
}
for(k in 1:SumNumNeigh){
        weights[k]<-1
        }
u[1:N]~car.normal(adj[],weights[],num[],tau.u)
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
        b4[i]~dnorm(0,tau.b4)
}
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
tau.b4~dgamma(0.1,0.0001)
}

############  modelo 16

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+
                alpha1*tempo[k]+alpha2*prop65[i,k]+alpha3*rend[i,k]+alpha4*csaude[i,k]+
                b1[i]*tempo[k]+b2[i]*prop65[i,k]+b3[i]*rend[i,k]+b4[i]*csaude[i,k]+u[i]
```

```
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
        b2[i]~dnorm(0,tau.b2)
        b3[i]~dnorm(0,tau.b3)
        b4[i]~dnorm(0,tau.b4)
}
for(i in 1:N){
        u[i]~dnorm(0,tau.u)
}
for(k in 1:T){
        tempo[k]<-k
}

alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
alpha2~dnorm(0,0.001)
alpha3~dnorm(0,0.001)
alpha4~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
tau.b2~dgamma(0.1,0.0001)
tau.b3~dgamma(0.1,0.0001)
tau.b4~dgamma(0.1,0.0001)
}


############  modelo 17

model{
for(i in 1:N){
        for(k in 1:T){
                Y[i,k]~dpois(mu[i,k])
                log(mu[i,k])<-log(E[i,k])+alpha0+
                alpha1*tempo[k]+
                b1[i]*tempo[k]+u[i]
        }
}
for(i in 1:N){
        b1[i]~dnorm(0,tau.b1)
}
for(k in 1:SumNumNeigh){
        weights[k]<-1
        }
u[1:N]~car.normal(adj[],weights[],num[],tau.u)
for(k in 1:T){
        tempo[k]<-k
}


alpha0~dnorm(0,0.001)
alpha1~dnorm(0,0.001)
tau.u~dgamma(0.1,0.0001)
tau.b1~dgamma(0.1,0.0001)
}
```

Appendix 16. Bayesian modelling analysis (R2OpenBUGS).

```
############################## modelos OpenBugs #################

library(R2OpenBUGS)
N<-18
T<-10
num<-c(3,3,3,3,5,5,4,1,4,4,3,3,4,6,4,1,4,6)
adj=c(
6,13,18,
7,8,15,
13,16,17,
9,17,18,
6,9,10,12,14,
1,5,9,10,18,
2,12,14,15,
2,
4,5,6,18,
5,6,11,14,
10,14,15,
5,7,14,
1,3,17,18,
5,7,10,11,12,15,
2,7,11,14,
3,
3,4,13,18,
1,4,6,9,13,17)
SumNumNeigh<-66


############### modelo 00
# todos os dados, apenas intercepto com efeitos aleatórios espacial

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL

data<-list("N","T","Y","E","num","adj","SumNumNeigh")
inits<-function(){
list(tau.u=1,alpha0=0,u=rep(0,18))
}
parameters <- c("alpha0","mu","u","tau.u")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun00.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)
# pD negativo

############### modelo 1
# todos os dados apenas intercepto e efeitos aleatórios
#(independentes e espacial)

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL

# espacial
data<-list("N","T","Y","E","num","adj","SumNumNeigh")
```

```
inits<-function(){
list(tau.u=10,tau.v=10,alpha0=0,u=rep(0,18),v=rep(0,18))
}
parameters <- c("alpha0","mu","u","tau.u","v","tau.v")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun01.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

# pD negativo

############### modelo 2
# todos os dados
# intercepto, efeito fixo tempo, e efeitos aleatórios

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL

# espacial
data<-list("N","T","Y","E","num","adj","SumNumNeigh")
inits<-function(){
list(tau.u=10,tau.v=10,alpha0=0,alpha1=0,u=rep(0,18),v=rep(0,18))
}
parameters <- c("alpha0","alpha1","mu","u","tau.u","v","tau.v")

result2 <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun02.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

## pD negativo

############### modelo 3
# todos os dados
# intercepto, efeito fixo tempo, e efeitos aleatórios espacial

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL


data<-list("N","T","Y","E")
inits<-function(){
list(tau.v=10,alpha0=0,alpha1=0,v=rep(0,18))
}
parameters <- c("alpha0","alpha1","mu","v","tau.v")

result3 <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun03.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 4

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
```

```
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL

# sem espacial
data<-list("N","T","Y","E","prop65")
inits<-function(){
list(tau.v=10,alpha0=0,alpha1=0,alpha2=0,v=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","mu","v","tau.v")

result4 <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                 model.file="modelJun04.txt", n.chains=1, n.iter=30000,
                 n.burnin=5000,n.thin=1,debug=T,save.history=F)

## pD>0


############### modelo 5
#

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL

data<-list("N","T","Y","E")
inits<-function(){
list(tau.u=10,tau.b1=10,alpha0=0,alpha1=0,u=rep(0,18),b1=rep(0,18))
}
parameters <- c("alpha0","alpha1","mu","u","tau.u","b1","tau.b1")

result5 <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                 model.file="modelJun05.txt", n.chains=1, n.iter=30000,
                 n.burnin=5000,n.thin=1,debug=T,save.history=F)

## pD>0

############### modelo 6

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL

data<-list("N","T","Y","E","prop65")
inits<-function(){
list(tau.u=10,tau.b1=10,alpha0=0,alpha1=0,alpha2=0,u=rep(0,18),b1=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","mu","u","tau.u","b1","tau.b1")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                 model.file="modelJun06.txt", n.chains=1, n.iter=30000,
                 n.burnin=5000,n.thin=1,debug=T,save.history=F)

#pD>0
```

```
############### modelo 7

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL

data<-list("N","T","Y","E","prop65")
inits<-function(){
list(tau.u=10,tau.b1=10,tau.b2=10,alpha0=0,alpha1=0,alpha2=0,u=rep(0,18),b1=rep(0,18),b2=rep(0,1
8))
}
parameters <- c("alpha0","alpha1","alpha2","mu","u","tau.u","b1","tau.b1","b2","tau.b2")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun07.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

# pD>0

############### modelo 8

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
mulhM65<-as.matrix(mulheresM65[,2:11])
colnames(mulhM65)<-NULL

data<-list("N","T","Y","E","prop65","mulhM65")
inits<-function(){
list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,alpha0=0,alpha1=0,alpha2=0,alpha3=0,
u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","alpha3","mu","u","tau.u",
          "b1","tau.b1","b2","tau.b2","b3","tau.b3")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun08.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 9

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
hospit<-as.matrix(hosprel[,2:11])
colnames(hospit)<-NULL

data<-list("N","T","Y","E","prop65","hospit")
```

```
inits<-function(){
list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,alpha0=0,alpha1=0,alpha2=0,alpha3=0,
u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","alpha3","mu","u","tau.u",
          "b1","tau.b1","b2","tau.b2","b3","tau.b3")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun09.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 10

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
hospit<-as.matrix(hosprel[,2:11])
colnames(hospit)<-NULL
rend<-as.matrix(rendimento[,2:11])
colnames(rend)<-NULL

data<-list("N","T","Y","E","prop65","hospit","rend")
inits<-function(){
list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,tau.b4=10,
alpha0=0,alpha1=0,alpha2=0,alpha3=0,alpha4=0,
u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18),b4=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","alpha3","alpha4","mu","u","tau.u",
          "b1","tau.b1","b2","tau.b2","b3","tau.b3","b4","tau.b4")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun10.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 11

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
hospit<-as.matrix(hosprel[,2:11])
colnames(hospit)<-NULL
rend<-as.matrix(rendimento[,2:11])
colnames(rend)<-NULL
csaude<-as.matrix(csauderel[,2:11])
colnames(csaude)<-NULL


data<-list("N","T","Y","E","prop65","hospit","rend","csaude")
inits<-function(){
list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,tau.b4=10,tau.b5=10,
alpha0=0,alpha1=0,alpha2=0,alpha3=0,alpha4=0,alpha5=0,
```

```
u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18),b4=rep(0,18),b5=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","alpha3","alpha4","alpha5",
          "mu","u","tau.u",
          "b1","tau.b1","b2","tau.b2","b3","tau.b3","b4","tau.b4","b5","tau.b5")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun11.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 12

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
hospit<-as.matrix(hosprel[,2:11])
colnames(hospit)<-NULL
rend<-as.matrix(rendimento[,2:11])
colnames(rend)<-NULL
csaude<-as.matrix(csauderel[,2:11])
colnames(csaude)<-NULL
mulhM65<-as.matrix(mulheresM65[,2:11])
colnames(mulhM65)<-NULL


data<-list("N","T","Y","E","prop65","mulhM65","hospit","rend","csaude")
inits<-function(){
list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,tau.b4=10,tau.b5=10,tau.b6=10,
alpha0=0,alpha1=0,alpha2=0,alpha3=0,alpha4=0,alpha5=0,alpha6=0,
u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18),b4=rep(0,18),b5=rep(0,18),b6=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","alpha3","alpha4","alpha5","alpha6",
          "mu","u","tau.u",
          "b1","tau.b1","b2","tau.b2","b3","tau.b3","b4","tau.b4","b5","tau.b5","tau.b6")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
                    model.file="modelJun12.txt", n.chains=1, n.iter=30000,
                    n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 14

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
rend<-as.matrix(rendimento[,2:11])
colnames(rend)<-NULL
csaude<-as.matrix(csauderel[,2:11])
colnames(csaude)<-NULL

data<-list("N","T","Y","E","prop65","rend","csaude")
inits<-function(){
```

```
  list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,tau.b4=10,
      u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18),b4=rep(0,18))
}
parameters <- c("mu","u","tau.u","b1","tau.b1","b2","tau.b2","b3","tau.b3","b4","tau.b4")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
              model.file="modelJun14.txt", n.chains=1, n.iter=30000,
              n.burnin=5000,n.thin=1,debug=T,save.history=F)


############### modelo 15

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
rend<-as.matrix(rendimento[,2:11])
colnames(rend)<-NULL
csaude<-as.matrix(csauderel[,2:11])
colnames(csaude)<-NULL

data<-list("N","T","Y","E","prop65","rend","csaude","num","adj","SumNumNeigh")
inits<-function(){
  list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,tau.b4=10,
      u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18),b4=rep(0,18))
}
parameters <- c("mu","u","tau.u","b1","tau.b1","b2","tau.b2","b3","tau.b3","b4","tau.b4")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
              model.file="modelJun15.txt", n.chains=1, n.iter=30000,
              n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 16

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL
Tempo<-c(1:T)
prop65<-as.matrix(propmais65[,2:11])
colnames(prop65)<-NULL
rend<-as.matrix(rendimento[,2:11])
colnames(rend)<-NULL
csaude<-as.matrix(csauderel[,2:11])
colnames(csaude)<-NULL

data<-list("N","T","Y","E","prop65","rend","csaude")
inits<-function(){
  list(tau.u=10,tau.b1=10,tau.b2=10,tau.b3=10,tau.b4=10,
      alpha0=0,alpha1=0,alpha2=0,alpha3=0,alpha4=0,
      u=rep(0,18),b1=rep(0,18),b2=rep(0,18),b3=rep(0,18),b4=rep(0,18))
}
parameters <- c("alpha0","alpha1","alpha2","alpha3","alpha4",
              "mu","u","tau.u",
              "b1","tau.b1","b2","tau.b2","b3","tau.b3","b4","tau.b4")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
```

```
            model.file="modelJun16.txt", n.chains=1, n.iter=30000,
            n.burnin=5000,n.thin=1,debug=T,save.history=F)

############### modelo 17

Y<-as.matrix(int[,2:11])
colnames(Y)<-NULL
E<-as.matrix(esperados[,2:11])
colnames(E)<-NULL


data<-list("N","T","Y","E","num","adj","SumNumNeigh")
inits<-function(){
  list(tau.u=10,tau.b1=10,alpha0=0,alpha1=0,
      u=rep(0,18),b1=rep(0,18))
}
parameters <- c("alpha0","alpha1","mu","u","tau.u","b1","tau.b1")

result <- bugs (data=data, inits=inits, parameters.to.save=parameters,
            model.file="modelJun17.txt", n.chains=1, n.iter=30000,
            n.burnin=5000,n.thin=1,debug=T,save.history=F)
```

Appendix 17. Scripts for GLMM fitting (R Statistics).

```
#### model 1
m1<-glmer(int~ano+offset(log(esp))+(1|distrito),data=data,family=poisson)
summary(m1)
print(m1)
coef(m1)
plot(data$int,exp(predict(m1)))
abline(0,1,col=2,lwd=2)
resm1<-resid(m1)
plot(resm1~data$distrito,las=2,xlab="",col=2:19)
abline(h=0,lty=2,col=2)


#### model 2
m2<-glmer(int~ano+offset(log(esp))+ (1+ano|distrito),data=data,family=poisson)
summary(m2)
coef(m2)
plot(data$int,exp(predict(m2)))
abline(0,1,col=2,lwd=2)
resm2<-resid(m2)
plot(resm2~data$distrito,las=2,xlab="",col=2:19)
abline(h=0,lty=2,col=2)

#### Compare m1 vs. m2
plot(compareFits(coef(m1)$distrito,coef(m2)$distrito),col=c(1,2))
anova(m1,m2)


#### model 3
m3<-glmer(int~ano+propmais65+offset(log(esp))+
(1+ano|distrito),data=data,family=poisson)
summary(m3)
coef(m3)
plot(data$int,exp(predict(m3)))
abline(0,1,col=2,lwd=2)
resm3<-resid(m3)
plot(resm3~data$distrito,las=2,xlab="",col=2:19)
abline(h=0,lty=2,col=2)

#### Compare m2 vs. m3
plot(compareFits(coef(m2a)$distrito,coef(m3)$distrito),col=c(1,2))
anova(m2,m3)


#### model 4
m4<-glmer(int~ano+propmais65+rend+offset(log(esp))+
(1+ano+propmais65|distrito),data=data,family=poisson)
coef(m4)
summary(m4)
plot(data$int,exp(predict(m4)))
abline(0,1,col=2,lwd=2)
resm4<-resid(m4)
plot(resm4~data$distrito,las=2,xlab="",col=2:19)
abline(h=0,lty=2,col=2)

#### compare m3 vs. m4
#plot(compareFits(coef(m3)$distrito,coef(m4)$distrito),col=c(1,2))
anova(m3a,m4)
```

```
length(coef(m3)$distrito)


#### model 4a
m4a<-glmer(int~ano+propmais65+rend+offset(log(esp))+
(1+ano+propmais65+rend|distrito),data=data,family=poisson)
coef(m4a)
summary(m4a)
anova(m4,m4a)


#### model 4b
m4b<-glmer(int~ano+propmais65+rend+acessocsaude+offset(log(esp))+
(1+ano+propmais65+rend|distrito),data=data,family=poisson)
coef(m4b)
summary(m4b)
anova(m4a,m4b)


#### model 4c
m4c<-glmer(int~ano+propmais65+rend+acessocsaude+offset(log(esp))+
(1+ano+propmais65+rend+acessocsaude|distrito),data=data,family=poisson)
coef(m4c)
summary(m4c)
anova(m4b,m4c)


#### model 4d
m4d<-glmer(int~ano+propmais65+rend+acessohosp+offset(log(esp))+
(1+ano+propmais65+rend|distrito),data=data,family=poisson)
coef(m4d)
summary(m4d)
anova(m4a,m4d)


#### model 4e
m4e<-glmer(int~ano+propmais65+rend+acessohosp+offset(log(esp))+
(1+ano+propmais65+rend+acessohosp|distrito),data=data,
family=poisson)
summary(m4e)
coef(m4e)
anova(m4a,m4e)
```

Appendix 18. Scripts for model 5 - GLMM without random effects for the region-related IV (R Statistics).

```
m5<-glmer(int~ano+propmais65+rend+acessohosp+offset(log(esp))+
(1+ano|distrito),data=data,family=poisson)
summary(m5)
```

Appendix 19. Scripts for model 6 – GEE (R Statistics).

```
m6<-geeglm(int~ano+propmais65+rend+acessohosp+offset(log(esp)),
data=data,id=distrito,family=poisson(link="log"),corstr="ar1")
summary(m6)
```

Appendix 20. . Observed and fitted number of hospital admissions (R Statistics).

```
pred_m4e<-matrix(exp(predict(m4e,data)),nrow=18,byrow=T)
pred_m5<-matrix(exp(predict(m5,data)),nrow=18,byrow=T)
pred_m6<-matrix(exp(predict(m6,data)),nrow=18,byrow=T)

par(mfrow=c(3,6))
for(i in 1:18){
        ymin<-max(0,min(min(observados[i,]),min(pred_m6[i,]))-100)
        ymax<-max(max(observados[i,]),max(pred_m6[i,]))
        plot(c(2003:2012),observados[i,],type="b",xlab="",ylab="",
        main=unique(data$distrito)[i],ylim=c(ymin,ymax))
        points(c(2003:2012),pred_m5[i,],col=3,type="b")
        points(c(2003:2012),pred_m4e[i,],col=2,type="b")
        points(c(2003:2012),pred_m6[i,],col=4,type="b")
}
```

Appendix 21. . Goodness of fit and residual analysis (R Statistics).

```
# fitted vs observed
prev<-predict(m4e,data)
plot(data$int,exp(prev),xlab="observed",ylab="fitted values", cex.lab=1.2)
abline(0,1)

res<-data$int-exp(predict(m4e,data))
plot(res)
plot(data$int,res)
plot(density(res))
hist(res)

PRes<-res/sqrt(exp(prev))
#Pearson residuals vs fitted
plot(exp(prev),PRes,xlab="fitted values",ylab="Pearson residuals",cex.lab=1.2)
abline(h=0,lty=2)

# Pearson residuals
plot(PRes~data$distrito,axes=FALSE,xlab="",ylab="Pearson Residuals")
abline(h=0)
abline(h=c(-3,3),lty=2)
box()
axis(1,las=2,labels=distritos, at=1:18)
axis(2)
```

Appendix 22. . Goodness of fit and residual analysis (R Statistics).

```
datasmall<-data[data$ano!=10,]
dim(data)
dim(datasmall)

mprev<-glmer(int~ano+propmais65+rend+acessohosp+offset(log(esp))+
(1+ano+propmais65+rend+acessohosp|distrito),data=datasmall,
family=poisson)
summary(mprev)

data2012<-data[data$ano==10,]
predict(mprev,newdata=data2012,re.form=~(1+ano+propmais65+rend+acessohosp|distrito))

pred2012<-exp(predict(mprev,newdata=data2012))
plot(data2012$int,pred2012,xlim=c(0,3000),ylim=c(0,3000),
    xlab="observed",ylab="predicted",cex.lab=1.3)
abline(0,1,col=2,lwd=2)
identify(data2012$int,pred2012,labels=data2012$distrito,cex=1.2)

erro2012<-data2012$int-pred2012
plot(erro2012,xlab="district id",ylab="prediction error",pch=19,ylim=c(-300,300),cex.lab=1.3)
abline(h=0,lty=2)
identify(erro2012,labels=data2012$distrito,cex=1.2)


# mean absolute deviation and sqrt mean squared error

sum(abs(erro2012))/18
sqrt(sum(erro2012^2)/18)
```

Appendix 23. Tables with the observed and predicted number of hospital admissions (R Statistics).

```
## Observed number of hospital admissions

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]  645  775  777  711  860  892  856  899  942   923
 [2,]  102  135  135  151  122  116  111  151  135   164
 [3,]  677  789  674  615  649  691  701  718  815   898
 [4,]  306  336  330  346  389  375  296  347  342   351
 [5,]  477  489  614  579  645  608  646  790  657   607
 [6,]  547  633  635  572  647  570  629  579  653   641
 [7,]   96   89   46  112  185  161  175  214  235   218
 [8,]  318  339  413  455  409  324  309  262  252   364
 [9,]  449  465  504  561  500  244  191  242  328   391
[10,]  606  691  367  394  415  717  747  703  844   916
[11,]  939 1230 1311 1491 1900 1874 2132 2194 2227  2466
[12,]  177  289  176  156  162  181  192  133  153    81
[13,] 1416 1580 1663 1854 2016 1917 1652 1548 2004  2125
[14,]  488  558  607  452  589  843  845  796  672   821
[15,]  405  302  366  445  635  644  759  717  788   905
[16,]  261  282  252  232  164  104  190  218  295   417
[17,]  269  306  224  355  424  406  489  494  469   573
[18,]  572  591  584  626  694  726  611  613  559   618


## Model 4e: prediction of the number of hospital admissions

       [,1]   [,2]   [,3] [,4] [,5] [,6] [,7] [,8] [,9]  [,10]
 [1,] 674.6  720.7  760.9  798  834  863  872  885  928  945.4
 [2,] 122.1  120.0  129.3  137  130  120  119  143  145  156.4
 [3,] 708.3  693.1  687.9  672  667  673  689  750  819  867.4
 [4,] 327.2  324.5  329.7  350  357  352  350  331  339  356.8
 [5,] 503.3  545.1  572.8  570  558  636  670  724  699  635.5
 [6,] 561.7  617.8  614.3  592  640  610  586  601  628  654.6
 [7,]  85.9   84.4   85.4  110  137  168  198  221  216  224.5
 [8,] 321.9  362.9  401.4  440  380  329  326  239  293  352.6
 [9,] 492.2  457.7  528.3  493  419  321  201  211  356  396.2
[10,] 563.9  505.5  519.8  496  560  622  634  749  915  834.7
[11,] 948.8 1186.4 1335.3 1531 1890 1870 2102 2185 2269 2445.8
[12,] 208.1  242.0  178.6  173  158  184  178  165  119   95.9
[13,] 1446.9 1647.5 1608.7 1823 1930 1800 1885 1580 1883 2169.9
[14,] 475.4  512.1  549.6  588  689  776  822  719  722  817.8
[15,] 317.8  361.8  419.5  489  571  645  753  721  799  889.5
[16,] 227.2  259.5  275.2  243  194  156  149  211  336  364.0
[17,] 260.7  282.9  293.7  348  407  434  442  444  507  589.6
[18,] 569.5  583.5  607.0  638  668  686  666  594  562  619.4
```

```
## Model 5: prediction of the number of hospital admissions

        [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
  [1,]  694.0  739  751  784  825  828  882  883  918   976
  [2,]  117.0  115  130  135  134  133  134  126  143   156
  [3,]  649.0  659  687  708  728  722  746  736  780   816
  [4,]  322.2  316  329  361  371  351  353  322  330   361
  [5,]  519.5  537  557  585  619  617  637  632  674   729
  [6,]  585.9  574  602  621  624  620  624  587  622   646
  [7,]   83.4   95  106  118  134  151  173  189  221   261
  [8,]  377.4  373  374  363  349  341  334  314  312   311
  [9,]  508.6  467  448  426  404  372  353  303  301   291
 [10,]  462.7  503  537  583  613  634  698  716  787   866
 [11,] 1111.7 1257 1365 1508 1632 1779 2000 2058 2380  2671
 [12,]  222.8  204  200  189  180  161  149  136  132   126
 [13,] 1544.6 1632 1646 1728 1787 1784 1843 1786 1946  2081
 [14,]  506.0  546  578  596  648  684  722  730  791   869
 [15,]  342.7  381  431  494  543  602  702  707  819   946
 [16,]  210.0  217  228  240  241  240  241  245  265   290
 [17,]  262.7  288  299  344  396  423  440  454  518   582
 [18,]  605.7  610  618  621  629  623  618  601  622   645

## Model 6: prediction of the number of hospital admissions

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
  [1,]  567  604  625  652  684  705  742  782  858   919
  [2,]  233  234  246  254  254  252  255  266  276   288
  [3,]  684  717  755  786  813  832  873  905  973  1035
  [4,]  245  253  269  274  288  290  305  324  338   376
  [5,]  333  347  361  376  393  397  411  428  451   477
  [6,]  503  533  555  570  596  604  619  655  711   760
  [7,]  220  231  241  253  262  268  280  285  292   311
  [8,]  409  429  448  472  487  503  527  551  599   647
  [9,]  302  311  328  342  353  360  364  373  404   425
 [10,]  461  487  508  528  545  563  585  621  669   715
 [11,] 1363 1443 1486 1534 1609 1634 1739 1834 1980  2124
 [12,]  210  216  230  228  228  228  232  247  260   270
 [13,] 1268 1341 1381 1462 1512 1544 1632 1710 1846  1999
 [14,]  512  540  563  575  606  622  652  679  729   776
 [15,]  602  623  649  677  698  721  771  772  831   890
 [16,]  305  318  334  350  364  369  378  401  432   462
 [17,]  278  292  296  323  338  349  355  370  396   416
 [18,]  465  486  508  529  546  563  585  613  654   690
```